

# Robust Bayesian compressed sensing over finite fields: asymptotic performance analysis

Wenjie Li, Francesca Bassi, and Michel Kieffer

## Abstract

This paper addresses the topic of robust Bayesian compressed sensing over finite fields. For stationary and ergodic sources, it provides asymptotic (with the size of the vector to estimate) necessary and sufficient conditions on the number of required measurements to achieve vanishing reconstruction error, in presence of sensing and communication noise. In all considered cases, the necessary and sufficient conditions asymptotically coincide. Conditions on the sparsity of the sensing matrix are established in presence of communication noise. Several previously published results are generalized and extended.

## Index Terms

Compressed sensing, finite fields, MAP estimation, asymptotic performance analysis, stationary and ergodic sources.

W. Li and M. Kieffer are with Laboratoire des Signaux et Systèmes, CNRS-Supelec-Univ Paris-Sud, Gif-sur-Yvette. M. Kieffer is partly supported by the Institut Universitaire de France. F. Bassi is with ESME-Sudria, Ivry-sur-Seine. This work was partly supported by European Network of Excellence project NEWCOM#.

## I. INTRODUCTION

Compressed sensing refers to the compression of a vector  $\boldsymbol{\theta} \in \mathbb{R}^N$ , obtained by acquiring linear measurements whose number  $M$  can be significantly smaller than the size of the vector itself. If  $\boldsymbol{\theta}$  is  $k$ -sparse with respect to some known basis, its almost surely exact reconstruction can be evaluated from the linear measurements using basis pursuit, for  $M$  as small as  $\mathcal{O}(k \log(N))$  [1], [2]. The same result holds true also for compressible vector  $\boldsymbol{\theta}$  [3], with reconstruction quality matching the one allowed by direct observation of the biggest  $k$  coefficients of  $\boldsymbol{\theta}$  in the transform domain. The major feature of compressed sensing is that the linear coefficients do not need to be adaptive with respect to the signal to be acquired, but can actually be random, provided that appropriate conditions on the global measurement matrix are satisfied [2], [4]. Moreover, compressed sensing is robust to the presence of noise in the measurements [4], [5].

Bayesian compressed sensing [6] refers to the same problem, considered in the statistical inference perspective. In particular, the vector to be compressed is now understood as a statistical source  $\boldsymbol{\Theta}$ , whose *a priori* distribution can induce sparsity or correlation between the symbols. This allows to redefine the reconstruction problem as an estimation problem, solvable using standard Bayesian techniques, *e.g.*, Maximum A Posteriori (MAP) estimation. In practical implementations, estimation from the linear measurements can be achieved exploiting statistical graphical models [7], *e.g.*, using belief propagation [8] as done in [9] for deterministic measurement matrices, and in [10] for random measurement matrices.

In this paper we address the topic of robust Bayesian compressed sensing over finite fields. The motivating example for considering this setting comes from the large and growing bulk of works devoted to data dissemination and collection in wireless sensor networks. Wireless sensor networks [11] are composed by autonomous nodes, with sensing capability of some physical phenomenon (*e.g.* temperature, or pressure). In order

to ensure ease of deployment and robustness, the communication between the nodes might need to be performed in absence of designated access points and of a hierarchical structure. At the network layer, dissemination of the measurements to all the nodes can be achieved using an asynchronous protocol based on random linear network coding (RLNC) [12]. In the protocol, each node in the network broadcasts a packet evaluated as the linear combination of the local measurement, and of the packets received from neighboring nodes. The linear coefficients are randomly chosen, and are sent in each packet header. Upon an appropriate number of communication rounds, each node has collected enough linearly independent combinations of the network measurements, and can perform decoding, by solving a system of linear equations. Due to the physical nature of the sensed phenomenon, and to the spatial distribution of the nodes in the network, correlation between the measurements at different nodes can be assumed, and exploited to perform decoding, as done in [13]–[17]. Recasting the problem in the Bayesian compressed sensing framework, the vector of the measurements at the nodes is interpreted as the compressible source  $\Theta$ , the network coding matrix as the sensing matrix, and the decoding at each node as the estimation operation.

Before transmission, all the measurements needs to be quantized. Quantization can be performed after the network encoding operation, as done in [17], where reconstruction on the real field is performed via  $\ell_1$ -norm minimization, or it can be done prior to the network encoding operation. For the latter choice, which is the target of this work, each quantization index is represented by an element of a finite field, from which the network coding coefficients (*i.e.*, the sensing coefficients in the compressed sensing framework) are chosen as well. This setting has been considered in [13], where exact MAP reconstruction is obtained solving a mixed-integer quadratic program, and in [14]–[16], where approximate MAP estimation is obtained using variants of the belief propagation algorithm.

The performance analysis of compressed sensing over finite fields has been addressed

in [15], [18], and [19]. The work in [19] does not consider Bayesian priors, and assumes a known sparsity level of  $\theta$ . Ideal decoding via  $\ell_0$ -norm minimization is assumed, and necessary and sufficient conditions for exact recovery are derived as functions of the size of the vector, its sparsity level, the number of measurements, and the sparsity of the sensing matrix. Numerical results show that the necessary and sufficient conditions coincide, as the size of  $\theta$  asymptotically increases. A Bayesian setting is considered in [18] and [15]. In [18] a prior distribution induces sparsity on the realization of  $\Theta$ , whose elements are assumed statistically independent. Using the method of types [20], the error exponent with respect to exact reconstruction using  $\ell_0$ -norm minimization is derived in absence of noise in the measurements, and the error exponent with respect to exact reconstruction using minimum-empirical entropy decoding is derived for noisy measurements. In [15] specific correlation patterns (pairwise, cluster) between the elements of  $\Theta$  are considered. Error exponents under MAP decoding are derived, only in case of absence of noise on the measurements.

The contribution of this work can be summarized as follows. We assume a Bayesian setting and we consider MAP decoding. Inspired by the work in [19], we aim to derive necessary and sufficient conditions for almost surely exact recovery of  $\theta$ , as its size asymptotically increases. We consider three classes of prior distributions on the source vector: *i*) the prior distribution is sparsity inducing, and the elements are statistically independent; *ii*) the vector  $\Theta$  is a Markov process; *iii*) the vector  $\Theta$  is an ergodic process. To the best of our knowledge, no analysis has been previously performed for the latter source model, which is quite general. We consider both sparse and dense sensing matrices. We consider two kinds of noises: *a*) the sensing noise, affecting the measurements prior to network coding (*i.e.*, prior to random projection acquisition in the compressed sensing framework); *b*) the communication noise, affecting the network coded packets (*i.e.*, the random projections in the compressed sensing framework). To the best of our knowledge, no analysis has been previously performed in presence of

both kinds of noise. Considering source model *i*), our results for the noiseless setting are compatible with the ones presented in [19]; in addition, we can formally prove the asymptotic convergence of necessary and sufficient conditions, and extend the bounds on the sparsity factor of the sensing matrix in presence of communication noise. The asymptotic analysis under MAP decoding, both for the noiseless case and in presence of communication noise *b*), are compatible with the results derived in [18], respectively under  $\ell_0$ -norm minimization decoding and under minimum-empirical entropy decoding. Error exponents for MAP decoding of correlated sources in the noiseless setting are compatible with the ones presented in [15], and are here extended to the case of arbitrary statistical structure, and presence of noise contamination both preceding and following the sensing operation.

The rest of the paper is organized as follows. Section II introduces the considered signal models in the context of data dissemination in a wireless sensor network. In Section III, we derive the necessary conditions for asymptotic almost surely exact recovery, both for the noiseless and noisy cases. Section IV describes the sufficient conditions and the error exponents under MAP decoding, for the noiseless case and in presence of communication noise only. In Section V, sensing noise is also taken into account. Section VI concludes the paper.

## II. SYSTEM MODEL AND PROBLEM SETUP

This section introduces the system model as well as various hypotheses on the sources and on the sensing and communication noises. In what follows, sans-serif font denotes random quantities while serif font denotes deterministic quantities. Matrices are in bold-face upper-case letters. A length  $n$  vector is in bold-face lower-case with a superscript  $n$ . Calligraphic font denotes set, except  $\mathcal{H}$ , which denotes the entropy rate. All logarithms are in base 2.

### A. The source model

Consider a wireless sensor network consisting of a set  $\mathcal{N}$  of  $N = |\mathcal{N}|$  sensors. The target physical phenomenon (*e.g.* the temperature) at the  $n$ -th sensor is represented by the random variable  $\Theta_n$ , taking values on a finite field  $\mathbb{F}_Q$  of size  $Q$ . Let  $\boldsymbol{\theta}^N$  be a realization of the random vector  $\boldsymbol{\Theta}^N = (\Theta_1, \dots, \Theta_N)$ , taking values in  $\mathbb{F}_Q^N$ . The vector  $\boldsymbol{\Theta}^N$  represents the source in the Bayesian compressed sensing framework. The probability mass function (pmf) associated with  $\boldsymbol{\Theta}^N$  is denoted by  $p(\boldsymbol{\theta}^N)$ , rather than  $p_{\boldsymbol{\Theta}^N}(\boldsymbol{\theta}^N)$ , for the sake of simplicity. In general, the analytic form of  $p(\boldsymbol{\theta}^N)$  depends on the characteristics of the observed phenomenon and of the topology of the sensor network. Here we consider three different models, defined as follows.

*SI: Sparse, Independent and identically distributed source.* Each element of the source vector  $\boldsymbol{\Theta}^N$  is independent and identically distributed (iid) with pmf  $p_{\Theta}(\cdot)$  and  $p_{\Theta}(0) > 0.5$ ,

$$p(\boldsymbol{\theta}^N) = \prod_{n=1}^N p_{\Theta}(\theta_n). \quad (1)$$

*StM: Stationary Markov model.* Let  $\boldsymbol{\theta}_n^{n+r-1} \in \mathbb{F}_Q^r$  denote the sequences  $(\theta_n, \dots, \theta_{n+r-1})$ . This is the stationary  $r$ -th order Markov model with  $r \in \mathbb{N}^+$  and  $1 \leq r \ll N$  and transition probability  $p(\theta_{n+r} \mid \boldsymbol{\theta}_n^{n+r-1})$ . The pmf of  $\boldsymbol{\Theta}^N$  may be written as

$$p(\boldsymbol{\theta}^N) = p(\boldsymbol{\theta}_1^r) \prod_{n=1}^{N-r} p(\theta_{n+r} \mid \boldsymbol{\theta}_n^{n+r-1}). \quad (2)$$

*GSE: General Stationary and Ergodic model.* This is the general case, without any further assumption apart from the ergodicity of the source.

### B. The sensing model

The considered system model is shown in Figure 1. Let  $\mathbf{x}_n \in \mathbb{F}_Q$  be the measurement of  $\Theta_n$  obtained by the  $n$ -th sensor. The random vector  $\mathbf{x}^N = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \in \mathbb{F}_Q^N$  is a copy of the source vector  $\boldsymbol{\Theta}^N$  corrupted by the *sensing noise*. The sensing noise models

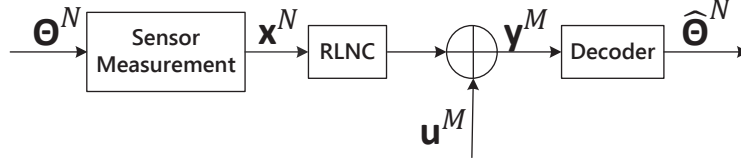


Figure 1. Block diagram for network compressive sensing model

the effect of imperfect measure acquisition at each sensor. It is described by the stationary transition probability  $p_{\mathbf{x}|\Theta}(x_n | \theta_n), \forall n$ . Remark that this implies that  $\mathbf{x}^N$  is stationary as long as  $\Theta^N$  is stationary. The local measurement  $x_n$  at node  $n$  is used to compute a packet via RLNC [12], which is then broadcast and received by the neighbours of  $n$ . Each node in the network can act as a sink, and attempt reconstruction of  $\Theta^N$ , after a number  $M \leq N$  of linear combinations has been received. The effects of RLNC at a sink node can be modeled as multiplying  $\mathbf{x}^N$  by a random matrix  $\mathbf{A} \in \mathbb{F}_Q^{M \times N}$ . We assume that some *communication noise*  $\mathbf{u}^M \in \mathbb{F}_Q^M$  affects the received packets, modeling the effects of transmission. Each entry of  $\mathbf{u}^M$  is iid with pmf  $p_u(\cdot)$ . The sink node is assumed to have received  $M$  packets, with the  $i$ -th packet carrying the coefficients  $\mathbf{A}_i$  and the result of linear combination  $y_i \in \mathbb{F}_Q$ , where  $\mathbf{A}_i$  is the  $i$ -th row of  $\mathbf{A}$  and  $y_i = \mathbf{A}_i \mathbf{x}^N + u_i$ , with all operations in  $\mathbb{F}_Q$ . The vector  $\mathbf{y}^M = (y_1, y_2, \dots, y_M)^t \in \mathbb{F}_Q^M$  can be then represented as

$$\mathbf{y}^M = \mathbf{A} \mathbf{x}^N + \mathbf{u}^M, \quad (3)$$

where the network coding matrix  $\mathbf{A}$  plays the role of the random sensing matrix in the compressed sensing setup. According to the presence of the sensing and communication noises, one obtains four types of noise models, namely Without Noise (WN), Noise in Communications (NC) only, Noise in the Sensing process (NS) only, and noise in both Communications and in the Sensing process (NCS). These models are summarized in Table I.

Table I  
CLASSIFICATION AND NOTATION BASED ON THE PRESENCE OF NOISE

		Communication Noise	
		absent	present
Sensing	absent	WN	NC
	present	NS	NCS

In general, the matrix  $\mathbf{A}$  is not necessarily of full rank, and it is assumed to be independent of  $\mathbf{x}^N$ . Two different assumptions about the structure of  $\mathbf{A}$  are considered here: (A1) the entries of  $\mathbf{A}$  are iid, uniformly distributed in  $\mathbb{F}_Q$ ; (A2) the entries of  $\mathbf{A}$  are iid, all non-zero elements of  $\mathbb{F}_Q$  are equiprobable. Both can be represented using the following model: for the entry  $A_{ij}$  of  $\mathbf{A}$ ,

$$\Pr(A_{ij} = q) = \begin{cases} 1 - \gamma & q = 0, \\ \gamma / (Q - 1) & q \in \mathbb{F}_Q \setminus \{0\}, \end{cases} \quad (4)$$

where  $\gamma$  is the sparsity factor,  $0 < \gamma < 1$ , and  $\mathbf{A}$  is sparse if  $\gamma < 0.5$ . We only assume that

$$0 < \gamma \leq 1 - Q^{-1}. \quad (5)$$

Notice that choosing  $\gamma < 1 - Q^{-1}$  corresponds to assumption (A2), while choosing  $\gamma = 1 - Q^{-1}$  corresponds to assumption (A1), since (4) becomes the uniform distribution.

In practice, sparse matrices are preferable. As the information of the sensing matrix is carried in the headers of packets [21], [22], the network coding overhead may be large if  $\mathbf{A}$  is dense and  $N$  is large. Moreover, as mentioned in [14], sparse matrices facilitate the convergence of the approximate belief propagation algorithm [8]. In practice, the structure of  $\mathbf{A}$  is strongly dependent on the structure of the network. For example, [15] assumes that only a subset of sensors  $\mathcal{S}_i \subset \mathcal{N}$  have participated in the  $i$ -th linear mixing. The content of the subsets  $\mathcal{S}_i$  depends on the location of each sensor and is designed



to minimize communication costs. In  $\mathbf{A}$ , coefficients associated to nodes belonging to  $\mathcal{S}_i$  follow a uniform distribution, while the others are null. This model, however, is not considered here, since we aim at a general asymptotic analysis, independent on the topology of the network.

### C. MAP Decoding

The sink node observes the realization  $\mathbf{y}^M$  and perfectly knows the realization  $\mathbf{A}$ , *e.g.*, from packet headers, see [21] and [22]. The maximum *a posteriori* estimate  $\hat{\boldsymbol{\theta}}^N$  of the realization of  $\boldsymbol{\Theta}^N$  is evaluated as

$$\hat{\boldsymbol{\theta}}^N = \arg \max_{\boldsymbol{\theta}^N \in \mathbb{F}_Q^N} p(\boldsymbol{\theta}^N | \mathbf{y}^M, \mathbf{A}), \quad (6)$$

where the *a posteriori* pmf is

$$\begin{aligned} p(\boldsymbol{\theta}^N | \mathbf{y}^M, \mathbf{A}) &\propto p(\boldsymbol{\theta}^N, \mathbf{y}^M, \mathbf{A}) \\ &= \sum_{\mathbf{x}^N \in \mathbb{F}_Q^N} \sum_{\mathbf{u}^M \in \mathbb{F}_Q^M} p(\boldsymbol{\theta}^N, \mathbf{x}^N, \mathbf{u}^M, \mathbf{y}^M, \mathbf{A}) \\ &= \sum_{\mathbf{x}^N \in \mathbb{F}_Q^N} \sum_{\mathbf{u}^M \in \mathbb{F}_Q^M} p(\boldsymbol{\theta}^N) p(\mathbf{x}^N | \boldsymbol{\theta}^N) p(\mathbf{u}^M) p(\mathbf{A}) p(\mathbf{y}^M | \mathbf{x}^N, \mathbf{u}^M, \mathbf{A}). \end{aligned} \quad (7)$$

Note that the conditional pmf  $p(\mathbf{y}^M | \mathbf{x}^N, \mathbf{u}^M, \mathbf{A})$  is an indicator function, *i.e.*,

$$p(\mathbf{y}^M | \mathbf{x}^N, \mathbf{u}^M, \mathbf{A}) = 1_{\mathbf{y}^M = \mathbf{A}\mathbf{x}^N + \mathbf{u}^M}. \quad (8)$$

An error event (decoding error) occurs when  $\hat{\boldsymbol{\theta}}^N \neq \boldsymbol{\theta}^N$ , with probability

$$P_e = \Pr \left\{ \hat{\boldsymbol{\Theta}}^N \neq \boldsymbol{\Theta}^N \right\}. \quad (9)$$

Our objective is to evaluate lower and upper bounds of (9) under MAP decoding, as functions of  $M$ ,  $N$ , and  $\gamma$ , for the various source and noise models previously introduced.

With these bounds, one can obtain necessary and sufficient conditions on the ratio  $M/N$  for asymptotic (with  $N \rightarrow \infty$ ) perfect recovery, *i.e.*, to obtain

$$\lim_{N \rightarrow \infty} P_e = 0. \quad (10)$$

### III. NECESSARY CONDITION FOR ASYMPTOTIC PERFECT RECOVERY

This section derives the necessary conditions for asymptotically ( $N \rightarrow \infty$ ) vanishing probability of decoding error. They only depend on the assumptions considered about the sensing and communication noises. We directly analyze the NCS case for the GSE source model. The results for this case can be easily adapted to the other cases. This work extends results obtained in [19] for the noiseless case (WN). Two situations are considered, depending on the value of the entropy rate

$$\mathcal{H}(\mathbf{x}) = \lim_{N \rightarrow \infty} \frac{1}{N} H(\mathbf{x}^N). \quad (11)$$

**Proposition 1** (Necessary condition for the NCS case). *Assume the presence of both communication and sensing noises and that  $\mathcal{H}(\mathbf{x}) > 0$ . Consider some arbitrary small  $\delta \in \mathbb{R}^+$ . For  $N \rightarrow \infty$ , the necessary conditions for  $P_e < \delta$  are*

$$\mathcal{H}(\Theta, \mathbf{x}) - \mathcal{H}(\mathbf{x}) < 3\varepsilon + \delta \log Q, \quad (12)$$

$$H(p_u) < \log Q, \quad (13)$$

and

$$\frac{M}{N} > \frac{\mathcal{H}(\Theta, \mathbf{x}) - (5\varepsilon + 2\delta \log Q)}{\log Q - H(p_u)}, \quad (14)$$

where  $\varepsilon \in \mathbb{R}^+$  is an arbitrary small constant.

**Corollary 1.** *Consider the same hypotheses as in Proposition 1 and assume now that  $\mathcal{H}(\mathbf{x}) = 0$ . Consider some arbitrary small  $\delta \in \mathbb{R}^+$ . For  $N \rightarrow \infty$ , the necessary condition for  $P_e < \delta$  is*

$$\mathcal{H}(\Theta, \mathbf{x}) < 3\varepsilon + \delta \log Q, \quad (15)$$

where  $\varepsilon \in \mathbb{R}^+$  is an arbitrary small constant.

In Proposition 1, (12) implies that for asymptotically exact recovery,  $p(\mathbf{x}^N | \Theta^N)$  should degenerate, almost surely, into a deterministic mapping. The condition (13) indicates that asymptotically exact recovery for non-deterministic sources is not possible in case of uniformly distributed communication noise. Finally, (14) indicates that the minimum number of required measurements depends both on the sensing and communication noises as well as on the distribution of  $\Theta$ . In particular, for a given source with entropy rate  $\mathcal{H}(\Theta)$ , the number of necessary measurements increases with the level of the sensing noise, determined by  $\mathcal{H}(\mathbf{x} | \Theta)$ . Similarly, the number of necessary measurements increases when the communication noise gets closer to uniformly distributed. The following proof is inspired by the work in [19], with both communication noise and sensing noise are considered here.

*Proof:* From the problem setup, one has the Markov chain

$$\Theta^N \leftrightarrow \mathbf{x}^N \leftrightarrow (\mathbf{y}^M, \mathbf{A}) \leftrightarrow \hat{\Theta}^N, \quad (16)$$

from which one deduces that

$$H(\Theta^N | \mathbf{x}^N) \leq H(\Theta^N | \hat{\Theta}^N), \quad (17)$$

and

$$H(\mathbf{x}^N | \mathbf{y}^M, \mathbf{A}) \leq H(\Theta^N | \hat{\Theta}^N). \quad (18)$$

Applying Fano's inequality [23, Sec. 2.10], one gets

$$\begin{aligned} H(\Theta^N | \hat{\Theta}^N) &\leq 1 + P_e \cdot \log(Q^N - 1) \\ &< 1 + NP_e \log Q, \end{aligned} \quad (19)$$

an upper bound of  $P_e$  is obtained combining (17) and (19),

$$P_e > \frac{H(\Theta^N, \mathbf{x}^N) - H(\mathbf{x}^N) - 1}{N \log Q}. \quad (20)$$

Since  $\Theta^N$  and  $\mathbf{x}^N$  are stationary and ergodic, for any  $\varepsilon > 0$ , there exists  $N_0 \in \mathbb{N}$  such that  $\forall N > N_0$ , one has

$$\begin{cases} \mathcal{H}(\Theta, \mathbf{x}) - \varepsilon < \frac{H(\Theta^N, \mathbf{x}^N)}{N} < \mathcal{H}(\Theta, \mathbf{x}) + \varepsilon, \\ \mathcal{H}(\mathbf{x}) - \varepsilon < \frac{H(\mathbf{x}^N)}{N} < \mathcal{H}(\mathbf{x}) + \varepsilon, \\ \varepsilon > \frac{1}{N}. \end{cases} \quad (21)$$

Hence for  $N > N_0$ , (20) can be rewritten as

$$P_e > \frac{\mathcal{H}(\Theta, \mathbf{x}) - \mathcal{H}(\mathbf{x}) - 3\varepsilon}{\log Q}. \quad (22)$$

For  $P_e < \delta$ , one deduces (12) from (22). For  $\delta$  and  $\varepsilon$  arbitrary small, (12) imposes  $\mathcal{H}(\Theta, \mathbf{x}) = \mathcal{H}(\mathbf{x})$ , meaning that  $\Theta$  should be deterministic knowing  $\mathbf{x}$ , almost surely.

From (18) and (19), one gets an other lower bound for  $P_e$

$$P_e > \frac{H(\mathbf{x}^N | \mathbf{y}^M, \mathbf{A}) - 1}{N \log Q}. \quad (23)$$

The conditional entropy  $H(\mathbf{x}^N | \mathbf{y}^M, \mathbf{A})$  can be bounded as

$$\begin{aligned} H(\mathbf{x}^N | \mathbf{y}^M, \mathbf{A}) &= H(\mathbf{x}^N) - I(\mathbf{x}^N; \mathbf{y}^M, \mathbf{A}) \\ &= H(\mathbf{x}^N) - (I(\mathbf{x}^N; \mathbf{A}) + I(\mathbf{x}^N; \mathbf{y}^M | \mathbf{A})) \\ &\stackrel{(a)}{=} H(\mathbf{x}^N) - (H(\mathbf{y}^M | \mathbf{A}) - H(\mathbf{y}^M | \mathbf{A}, \mathbf{x}^N)) \\ &\stackrel{(b)}{\geq} H(\mathbf{x}^N) - M \cdot \log Q + H(\mathbf{y}^M | \mathbf{A}, \mathbf{x}^N) \\ &\stackrel{(c)}{=} H(\mathbf{x}^N) - M \cdot \log Q + MH(p_u), \end{aligned} \quad (24)$$

where (a) follows from the assumption that  $\mathbf{x}^N$  and  $\mathbf{A}$  are independent, (b) comes from  $H(\mathbf{y}^M | \mathbf{A}) \leq H(\mathbf{y}^M) \leq \log |\mathbb{F}_Q^M| = M \log Q$ , and (c) is because

$$H(\mathbf{y}^M | \mathbf{A}, \mathbf{x}^N) = H(\mathbf{A}\mathbf{x}^N + \mathbf{u}^M | \mathbf{A}, \mathbf{x}^N) = H(\mathbf{u}^M) = MH(p_u). \quad (25)$$

Using (23) and (24), a second necessary condition for  $P_e < \delta$  is

$$\frac{H(\mathbf{x}^N) - M(\log Q - H(p_u)) - 1}{N \log Q} < \delta. \quad (26)$$

For  $N > N_0$ , using (21) in (26) yields

$$\frac{\mathcal{H}(\mathbf{x}) - \frac{M}{N}(\log Q - H(p_u)) - 2\varepsilon}{\log Q} < \delta. \quad (27)$$

Now consider two cases. In the first case, the communication noise is assumed uniformly distributed, *i.e.*,

$$H(p_u) = \log Q, \quad (28)$$

the condition (27) becomes

$$\mathcal{H}(\mathbf{x}) < \delta \log Q + 2\varepsilon. \quad (29)$$

As  $\delta$  can be made arbitrary small, (29) imposes that, for uniform communication noise, asymptotically vanishing probability of error is possible only if  $\mathcal{H}(\mathbf{x})$  is arbitrary close to zero. For non-degenerate cases, *i.e.*,  $\mathcal{H}(\mathbf{x}) > 0$ , one obtains the necessary condition (13). In this second case, a lower bound of the compression ratio  $M/N$  is obtained immediately from (27),

$$\frac{M}{N} > \frac{\mathcal{H}(\mathbf{x}) - (2\varepsilon + \delta \log Q)}{\log Q - H(p_u)}. \quad (30)$$

We can represent the condition (30) in terms of the joint entropy rate  $\mathcal{H}(\Theta, \mathbf{x})$  by applying (12). Then, one gets (14) and Proposition 1 is proved.

Consider now  $\mathcal{H}(\mathbf{x}) = 0$ , then (27) holds for any value of  $M/N$ , and for any  $H(p_u) \leq \log Q$ , since the left side of (27) is always negative. Hence, (12) is the only necessary condition for this case. Corollary 1 is also proved. ■

With the results of the NCS noise model, one may derive the necessary conditions for the other models. If no sensing noise is considered, *i.e.*,  $\mathbf{x}^N = \Theta^N$ , one has  $H(\Theta^N | \mathbf{x}^N) = H(\mathbf{x}^N | \Theta^N) = 0$  and  $H(\Theta^N, \mathbf{x}^N) = H(\Theta^N)$ . If communication noise is absent, *i.e.*,  $\mathbf{u}^M = \mathbf{0}$ ,  $H(\mathbf{u}^M) = 0$ . The necessary conditions for asymptotically ( $N \rightarrow \infty$ ) vanishing probability of decoding error for each case are listed in Table 2.

Table II  
NECESSARY CONDITIONS FOR ASYMPTOTIC PERFECT RECOVERY IN NOISELESS AND NOISY CASES

Case	Necessary Condition ( $\mathcal{H}(\mathbf{x}) > 0$ )
WN	$\frac{M}{N} > \frac{\mathcal{H}(\Theta)}{\log Q}$ , already obtained in [19],
NC	$\frac{M}{N} > \frac{\mathcal{H}(\Theta)}{\log Q - H(p_u)}$ and $H(p_u) < \log Q$ ,
NS	$\frac{M}{N} > \frac{\mathcal{H}(\Theta, \mathbf{x})}{\log Q}$ and $\mathcal{H}(\Theta   \mathbf{x}) = 0$ ,
NCS	$\frac{M}{N} > \frac{\mathcal{H}(\Theta, \mathbf{x})}{\log Q - H(p_u)}$ and $H(p_u) < \log Q$ and $\mathcal{H}(\Theta   \mathbf{x}) = 0$ .

#### IV. SUFFICIENT CONDITION IN ABSENCE OF SENSING NOISE

This section provides an upper bound of the error probability for the MAP estimation problem in absence of sensing noise (the WN and NC cases). These two cases are considered simultaneously because their proofs are similar. When the channel noise vanishes, the NC case boils down to the WN case.

##### A. Upper Bound of the Error Probability

**Proposition 2** (Upper bound of  $P_e$ , WN and NC cases). *Under MAP decoding, the asymptotic ( $N \rightarrow \infty$ ) probability of error in absence of sensing noise can be upper bounded as*

$$P_e \leq P_1(\alpha) + P_2(\alpha) + 2\varepsilon, \quad (31)$$

where  $\varepsilon \in \mathbb{R}^+$  is an arbitrarily small constant.  $P_1(\alpha)$  and  $P_2(\alpha)$  are defined as

$$P_1(\alpha) = 2^{-N \left( -\frac{M}{N} (H(p_u) + \log(1-\gamma) + \varepsilon) - H_2(\alpha) - \alpha \log(Q-1) - \frac{\log(\alpha N)}{N} \right)}, \quad (32)$$

and

$$P_2(\alpha) = 2^{-N \left( -\mathcal{H}(\Theta) - \frac{M}{N} \left( H(p_u) + \log \left( Q^{-1} + \left( 1 - \frac{\gamma}{1-Q^{-1}} \right)^{\lceil \alpha N \rceil} (1-Q^{-1}) \right) + \varepsilon \right) - \varepsilon \right)}, \quad (33)$$

with  $\alpha \in \mathbb{R}^+$  and  $\alpha < 0.5$ .

*Proof:* The proof consists of two parts. First we define the error event, and then we analyze the probability of error.

Since no sensing noise is considered, we have  $\mathbf{x}^N = \boldsymbol{\Theta}^N$  throughout this section. The *a posteriori* pmf (7) becomes

$$p(\boldsymbol{\theta}^N | \mathbf{y}^M, \mathbf{A}) \propto \sum_{\mathbf{u}^M \in \mathbb{F}_Q^M} p(\boldsymbol{\theta}^N) p(\mathbf{u}^M) p(\mathbf{A}) 1_{\mathbf{y}^M = \mathbf{A}\boldsymbol{\theta}^N + \mathbf{u}^M}. \quad (34)$$

Suppose that  $\boldsymbol{\theta}^N$  (given but unknown) is the true state vector and consider that  $\mathbf{A}$  has been generated randomly. At the sink,  $\mathbf{A}$  and  $\mathbf{y}^M$  are known. With MAP decoding, the reconstruction  $\hat{\boldsymbol{\theta}}^N$  in (6) is

$$\hat{\boldsymbol{\theta}}^N = \arg \max_{\boldsymbol{\theta}^N \in \mathbb{F}_Q^N} \sum_{\mathbf{u}^M \in \mathbb{F}_Q^M} p(\boldsymbol{\theta}^N) p(\mathbf{u}^M) p(\mathbf{A}) 1_{\mathbf{y}^M = \mathbf{A}\boldsymbol{\theta}^N + \mathbf{u}^M}. \quad (35)$$

A decoding error happens if there exists a vector  $\boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\}$  such that

$$\sum_{\mathbf{v}^M \in \mathbb{F}_Q^M} p(\boldsymbol{\varphi}^N) p(\mathbf{v}^M) 1_{\mathbf{y}^M = \mathbf{A}\boldsymbol{\varphi}^N + \mathbf{v}^M} \geq \sum_{\mathbf{u}^M \in \mathbb{F}_Q^M} p(\boldsymbol{\theta}^N) p(\mathbf{u}^M) 1_{\mathbf{y}^M = \mathbf{A}\boldsymbol{\theta}^N + \mathbf{u}^M}. \quad (36)$$

For fixed  $\mathbf{y}^M$ ,  $\mathbf{A}$ , and  $\boldsymbol{\theta}^N$ , there is exactly one vector  $\mathbf{u}^M$  such that  $\mathbf{u}^M = \mathbf{y}^M - \mathbf{A}\boldsymbol{\theta}^N$ . Hence the right side of (36) can be represented as  $p_{\boldsymbol{\theta}^N}(\boldsymbol{\theta}^N) p_{\mathbf{u}^M}(\mathbf{y}^M - \mathbf{A}\boldsymbol{\theta}^N)$ . The subscripts for the pmfs are introduced to avoid any ambiguity of notations. Then (36) is equivalent to

$$p_{\boldsymbol{\theta}^N}(\boldsymbol{\varphi}^N) p_{\mathbf{u}^M}(\mathbf{y}^M - \mathbf{A}\boldsymbol{\varphi}^N) \geq p_{\boldsymbol{\theta}^N}(\boldsymbol{\theta}^N) p_{\mathbf{u}^M}(\mathbf{y}^M - \mathbf{A}\boldsymbol{\theta}^N). \quad (37)$$

An alternative way to state the error event can be: For a given realization  $\boldsymbol{\Theta}^N = \boldsymbol{\theta}^N$ , which implies the realization  $\mathbf{u}^M = \mathbf{u}^M = \mathbf{y}^M - \mathbf{A}\boldsymbol{\theta}^N$ , there exists a pair  $(\boldsymbol{\varphi}^N, \mathbf{v}^M) \in$

$\mathbb{F}_Q^N \times \mathbb{F}_Q^M$  such that

$$\begin{cases} \boldsymbol{\varphi}^N \neq \boldsymbol{\theta}^N, \\ \mathbf{A}\boldsymbol{\varphi}^N + \mathbf{v}^M = \mathbf{y}^M = \mathbf{A}\boldsymbol{\theta}^N + \mathbf{u}^M, \\ p(\boldsymbol{\varphi}^N) p(\mathbf{v}^M) \geq p(\boldsymbol{\theta}^N) p(\mathbf{u}^M). \end{cases} \quad (38)$$

From conditions (38), one concludes that the MAP decoder is equivalent to the maximum  $Q$ -probability decoder [24] in the NC case.

An upper bound of the error probability is now derived. For a fixed  $\boldsymbol{\theta}^N$  and  $\mathbf{u}^M$ , the conditional error probability is denoted by  $\Pr\{\text{error} \mid \boldsymbol{\theta}^N, \mathbf{u}^M\}$ . The average error probability is

$$P_e = \sum_{\boldsymbol{\theta}^N \in \mathbb{F}_Q^N} \sum_{\mathbf{u}^M \in \mathbb{F}_Q^M} p(\boldsymbol{\theta}^N, \mathbf{u}^M) \Pr\{\text{error} \mid \boldsymbol{\theta}^N, \mathbf{u}^M\}. \quad (39)$$

Weak typicality is instrumental in the following proofs. The notations of [25, Definition 4.2] are extended to stationary and ergodic sources. For any positive real number  $\varepsilon$  and some integer  $N > 0$ , the weakly typical set  $\mathcal{A}_{[\Theta]\varepsilon}^N \subset \mathbb{F}_Q^N$  for a stationary and ergodic source  $\Theta^N$  is the set of vectors  $\boldsymbol{\theta}^N \in \mathbb{F}_Q^N$  satisfying

$$\left| -\frac{1}{N} \log p(\boldsymbol{\theta}^N) - \mathcal{H}(\Theta) \right| \leq \varepsilon, \quad (40)$$

where  $\mathcal{H}(\Theta)$  is the entropy rate of the source. Similarly, for the noise vector  $\mathbf{u}^M$ , define

$$\mathcal{A}_{[\mathbf{u}]\varepsilon}^M = \left\{ \mathbf{u}^M \in \mathbb{F}_Q^M : \left| -\frac{1}{M} \log p(\mathbf{u}^M) - H(p_{\mathbf{u}}) \right| \leq \varepsilon \right\}. \quad (41)$$

Recall that the entries of  $\mathbf{u}^M$  are uncorrelated, so  $\mathcal{H}(\mathbf{u}) = H(p_{\mathbf{u}})$ . Thanks to Shannon-McMillan-Breiman theorem [23, Sec. 16.8], the pmf of the general stationary and ergodic source converges. In other words, for any  $\varepsilon > 0$ , there exists  $N_\varepsilon$  and  $M_\varepsilon$  such that for all  $N > N_\varepsilon$  and  $M > M_\varepsilon$ ,

$$\Pr \left\{ \left| -\frac{1}{N} \log p(\Theta^N) - \mathcal{H}(\Theta) \right| \leq \varepsilon \right\} \geq 1 - \varepsilon, \quad (42)$$



and

$$\Pr \left\{ \left| -\frac{1}{M} \log p(\mathbf{u}^M) - \mathcal{H}(p_u) \right| \leq \varepsilon \right\} \geq 1 - \varepsilon. \quad (43)$$

We can make  $\varepsilon$  arbitrary close to zero as  $N \rightarrow \infty$  and  $M \rightarrow \infty$ . A sandwich proof of this theorem is proposed in [23, Sec. 16.8]. For the sparse and uncorrelated source as defined in (1),  $\mathcal{H}(\Theta)$  is equal to  $H(p_\Theta)$ , the entropy of a single source. The entropy rate of the StM source is the conditional entropy  $H(\Theta_{n+r} \mid \Theta_n^{n+r-1})$ .

From (42) and (43), one has  $\Pr \left\{ \Theta^N \in \mathcal{A}_{[\Theta]\varepsilon}^N \right\} \geq 1 - \varepsilon$  and  $\Pr \left\{ \mathbf{u}^M \in \mathcal{A}_{[\mathbf{u}]\varepsilon}^M \right\} \geq 1 - \varepsilon$  for  $N > N_\varepsilon$  and  $M > M_\varepsilon$ . It implies that, for  $N$  and  $M$  sufficiently large,  $\Theta^N$  and  $\mathbf{u}^M$  belong to the weakly typical set  $\mathcal{A}_{[\Theta]\varepsilon}^N$  and  $\mathcal{A}_{[\mathbf{u}]\varepsilon}^M$ , almost surely. With respect to the typicality,  $\mathbb{F}_Q^N \times \mathbb{F}_Q^M$  can be divided into two parts. Define the sets  $\mathcal{U}$  and  $\mathcal{U}^c$  for the pair of vectors  $(\theta^N, \mathbf{u}^M)$ , such that  $\mathcal{U} \cup \mathcal{U}^c = \mathbb{F}_Q^N \times \mathbb{F}_Q^M$  and

$$\mathcal{U} = \left\{ \theta^N \in \mathbb{F}_Q^N, \mathbf{u}^M \in \mathbb{F}_Q^M : \theta^N \in \mathcal{A}_{[\Theta]\varepsilon}^N \text{ and } \mathbf{u}^M \in \mathcal{A}_{[\mathbf{u}]\varepsilon}^M \right\}, \quad (44)$$

$$\mathcal{U}^c = \left\{ \theta^N \in \mathbb{F}_Q^N, \mathbf{u}^M \in \mathbb{F}_Q^M : \theta^N \notin \mathcal{A}_{[\Theta]\varepsilon}^N \text{ or } \mathbf{u}^M \notin \mathcal{A}_{[\mathbf{u}]\varepsilon}^M \right\}. \quad (45)$$

$\mathcal{U}$  is the joint typical set for  $(\theta^N, \mathbf{u}^M)$ , due to the independence of  $\Theta^N$  and  $\mathbf{u}^M$ . The error probability can be bounded as

$$\begin{aligned} P_e &= \sum_{(\theta^N, \mathbf{u}^M) \in \mathcal{U}} p(\theta^N) p(\mathbf{u}^M) \cdot \Pr \{ \text{error} \mid \theta^N, \mathbf{u}^M \} \\ &\quad + \sum_{(\theta^N, \mathbf{u}^M) \in \mathcal{U}^c} p(\theta^N) p(\mathbf{u}^M) \cdot \Pr \{ \text{error} \mid \theta^N, \mathbf{u}^M \} \\ &\stackrel{(a)}{\leq} \sum_{(\theta^N, \mathbf{u}^M) \in \mathcal{U}} p(\theta^N) p(\mathbf{u}^M) \cdot \Pr \{ \text{error} \mid \theta^N, \mathbf{u}^M \} + \sum_{(\theta^N, \mathbf{u}^M) \in \mathcal{U}^c} p(\theta^N) p(\mathbf{u}^M) \\ &\stackrel{(b)}{\leq} \sum_{(\theta^N, \mathbf{u}^M) \in \mathcal{U}} p(\theta^N) p(\mathbf{u}^M) \cdot \Pr \{ \text{error} \mid \theta^N, \mathbf{u}^M \} + 2\varepsilon, \end{aligned} \quad (46)$$

where (a) comes from  $\Pr(\text{error} \mid \boldsymbol{\theta}^N, \mathbf{u}^M) \leq 1$  and (b) follows from the fact that

$$\begin{aligned}
 \sum_{(\boldsymbol{\theta}^N, \mathbf{u}^M) \in \mathcal{U}^c} p(\boldsymbol{\theta}^N) p(\mathbf{u}^M) &= 1 - \sum_{(\boldsymbol{\theta}^N, \mathbf{u}^M) \in \mathcal{U}} p(\boldsymbol{\theta}^N) p(\mathbf{u}^M) \\
 &= 1 - \sum_{\boldsymbol{\theta}^N \in \mathcal{A}_{[\boldsymbol{\theta}]}^N} p(\boldsymbol{\theta}^N) \sum_{\mathbf{u}^M \in \mathcal{A}_{[\mathbf{u}]}^M} p(\mathbf{u}^M) \\
 &\leq 1 - (1 - \varepsilon)(1 - \varepsilon) \leq 2\varepsilon.
 \end{aligned} \tag{47}$$

Since  $\mathbf{A}$  is generated randomly, define the random event

$$\mathcal{E}(\boldsymbol{\theta}^N, \mathbf{u}^M; \boldsymbol{\varphi}^N, \mathbf{v}^M) = \{\mathbf{A}\boldsymbol{\theta}^N + \mathbf{u}^M = \mathbf{A}\boldsymbol{\varphi}^N + \mathbf{v}^M\}, \tag{48}$$

where  $(\boldsymbol{\theta}^N, \mathbf{u}^M)$  is the realization of the environment state, and  $(\boldsymbol{\varphi}^N, \mathbf{v}^M)$  is the potential reconstruction result. Conditioned on  $(\boldsymbol{\theta}^N, \mathbf{u}^M)$ ,  $\Pr\{\text{error} \mid \boldsymbol{\theta}^N, \mathbf{u}^M\}$  is in fact the probability of the union of the events  $\mathcal{E}(\boldsymbol{\theta}^N, \mathbf{u}^M; \boldsymbol{\varphi}^N, \mathbf{v}^M)$  with all the parameter pairs  $(\boldsymbol{\varphi}^N, \mathbf{v}^M) \in \mathbb{F}_Q^N \times \mathbb{F}_Q^M$  such that  $\boldsymbol{\varphi}^N \neq \boldsymbol{\theta}^N$  and  $p(\boldsymbol{\varphi}^N)p(\mathbf{v}^M) \geq p(\boldsymbol{\theta}^N)p(\mathbf{u}^M)$ , see (38). The conditional error probability can then be rewritten as

$$\Pr\{\text{error} \mid \boldsymbol{\theta}^N, \mathbf{u}^M\} = \Pr\left\{ \bigcup_{\substack{\boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\}, \mathbf{v}^M \in \mathbb{F}_Q^M: \\ p(\boldsymbol{\varphi}^N)p(\mathbf{v}^M) \geq p(\boldsymbol{\theta}^N)p(\mathbf{u}^M)}} \mathcal{E}(\boldsymbol{\theta}^N, \mathbf{u}^M; \boldsymbol{\varphi}^N, \mathbf{v}^M) \right\}. \tag{49}$$

Introducing (49) in (46) and applying the union bound yields

$$\begin{aligned}
 P_e &\leq \sum_{(\boldsymbol{\theta}^N, \mathbf{u}^M) \in \mathcal{U}} p(\boldsymbol{\theta}^N) p(\mathbf{u}^M) \sum_{\substack{\boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\}, \mathbf{v}^M \in \mathbb{F}_Q^M: \\ p(\boldsymbol{\varphi}^N)p(\mathbf{v}^M) \geq p(\boldsymbol{\theta}^N)p(\mathbf{u}^M)}} \Pr\{\mathcal{E}(\boldsymbol{\theta}^N, \mathbf{u}^M; \boldsymbol{\varphi}^N, \mathbf{v}^M)\} + 2\varepsilon \\
 &= \sum_{(\boldsymbol{\theta}^N, \mathbf{u}^M) \in \mathcal{U}} p(\boldsymbol{\theta}^N) p(\mathbf{u}^M) \\
 &\quad \cdot \sum_{\substack{\boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\} \\ \mathbf{v}^M \in \mathbb{F}_Q^M}} \Phi(\boldsymbol{\theta}^N, \mathbf{u}^M; \boldsymbol{\varphi}^N, \mathbf{v}^M) \Pr\{\mathcal{E}(\boldsymbol{\theta}^N, \mathbf{u}^M; \boldsymbol{\varphi}^N, \mathbf{v}^M)\} + 2\varepsilon,
 \end{aligned} \tag{50}$$

where

$$\Phi(\boldsymbol{\theta}^N, \mathbf{u}^M; \boldsymbol{\varphi}^N, \mathbf{v}^M) = \begin{cases} 1 & \text{if } p(\boldsymbol{\varphi}^N) p(\mathbf{v}^M) \geq p(\boldsymbol{\theta}^N) p(\mathbf{u}^M), \\ 0 & \text{if } p(\boldsymbol{\varphi}^N) p(\mathbf{v}^M) < p(\boldsymbol{\theta}^N) p(\mathbf{u}^M). \end{cases} \quad (51)$$

Now consider the following lemma.

**Lemma 1** (Upper bound of  $\Phi$ ). *Consider some  $s \in \mathbb{R}^+$  with  $s \leq 1$ . For any  $\boldsymbol{\theta}^N, \boldsymbol{\varphi}^N$  in  $\mathbb{F}_Q^N$  and  $\mathbf{u}^M, \mathbf{v}^M$  in  $\mathbb{F}_Q^M$ , the following inequality holds,*

$$\Phi(\boldsymbol{\theta}^N, \boldsymbol{\varphi}^N, \mathbf{u}^M, \mathbf{v}^M) \leq \left( \frac{p(\boldsymbol{\varphi}^N) p(\mathbf{v}^M)}{p(\boldsymbol{\theta}^N) p(\mathbf{u}^M)} \right)^s. \quad (52)$$

Lemma 1 is a part of Gallager's derivation of error exponents in [26, Sec. 5.6].

Introducing (52) with  $s = 1$  into (50), one gets

$$P_e \leq \sum_{(\boldsymbol{\theta}^N, \mathbf{u}^M) \in \mathcal{U}} \sum_{\substack{\boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\} \\ \mathbf{v}^M \in \mathbb{F}_Q^M}} p(\boldsymbol{\varphi}^N) p(\mathbf{v}^M) \Pr\{\mathcal{E}(\boldsymbol{\theta}^N, \mathbf{u}^M; \boldsymbol{\varphi}^N, \mathbf{v}^M)\} + 2\varepsilon. \quad (53)$$

In (53),

$$\Pr\{\mathcal{E}(\boldsymbol{\theta}^N, \mathbf{u}^M; \boldsymbol{\varphi}^N, \mathbf{v}^M)\} = \Pr\{\mathbf{A}\boldsymbol{\mu}^N = \mathbf{s}^M \mid \boldsymbol{\mu}^N \neq \mathbf{0}, \mathbf{s}^M\} \quad (54)$$

with  $\boldsymbol{\mu}^N = \boldsymbol{\varphi}^N - \boldsymbol{\theta}^N \in \mathbb{F}_Q^N \setminus \{\mathbf{0}\}$ , and  $\mathbf{s}^M = \mathbf{v}^M - \mathbf{u}^M \in \mathbb{F}_Q^M$ . This probability depends on the sparsity of  $\boldsymbol{\mu}^N$  and of  $\mathbf{s}^M$ , let  $d_1 = \|\boldsymbol{\mu}^N\|_0$  and  $d_2 = \|\mathbf{s}^M\|_0$ . Both  $d_1$  and  $d_2$  are integers such that  $1 \leq d_1 \leq N$  and  $0 \leq d_2 \leq M$ . Define the multivariable function

$$f(d_1, d_2; \gamma, Q, M) = \Pr\{\mathbf{A}\boldsymbol{\mu}^N = \mathbf{s}^M \mid \|\boldsymbol{\mu}^N\|_0 = d_1, \|\mathbf{s}^M\|_0 = d_2\}, \quad (55)$$

where  $\gamma$ ,  $Q$ , and  $M$  are the parameters of the random matrix  $\mathbf{A}$ .

$$\Pr\{\mathbf{A}\boldsymbol{\mu}^N = \mathbf{0} \mid \boldsymbol{\mu}^N \neq \mathbf{0}\} = f(d_1, 0; \gamma, Q, M)$$

has been evaluated in [27] and [19]. We provide a simple extension of this result for  $d_2 \neq 0$ .

**Lemma 2** (Properties of  $f(d_1, d_2; \gamma, Q, M)$ ). *The function  $f(d_1, d_2; \gamma, Q, M)$ , defined in (55), is non-increasing in  $d_2$  for a given  $d_1$  and*

$$f(d_1, d_2; \gamma, Q, M) \leq f(d_1, 0; \gamma, Q, M) = \left( Q^{-1} + \left( 1 - \frac{\gamma}{1 - Q^{-1}} \right)^{d_1} (1 - Q^{-1}) \right)^M. \quad (56)$$

*Moreover  $f(d_1, 0; \gamma, Q, M)$  is non-increasing in  $d_1$  and*

$$f(d_1, 0; \gamma, Q, M) \leq f(1, 0; \gamma, Q, M) = (1 - \gamma)^M. \quad (57)$$

*If  $\gamma = 1 - Q^{-1}$ , which corresponds to a uniformly distributed sensing matrix,*

$$f(d_1, d_2; \gamma, Q, M) = Q^{-M} \quad (58)$$

*is constant.*

See Appendix A for the proof details. Using Lemma 2, (53) can be expressed as

$$\begin{aligned} P_e &\stackrel{(a)}{\leq} \sum_{d_1=1}^N \sum_{d_2=0}^M \sum_{\substack{(\boldsymbol{\theta}^N, \mathbf{u}^M) \in \mathcal{U} \\ \boldsymbol{\varphi}^N \in \mathbb{F}_Q^N: \|\boldsymbol{\varphi}^N - \boldsymbol{\theta}^N\|_0 = d_1 \\ \mathbf{v}^M \in \mathbb{F}_Q^M: \|\mathbf{u}^M - \mathbf{v}^M\|_0 = d_2}} p(\boldsymbol{\varphi}^N) p(\mathbf{v}^M) f(d_1, d_2; \gamma, Q, M) + 2\varepsilon \\ &\stackrel{(b)}{\leq} \sum_{d_1=1}^N \sum_{\substack{(\boldsymbol{\theta}^N, \mathbf{u}^M) \in \mathcal{U} \\ \boldsymbol{\varphi}^N \in \mathbb{F}_Q^N: \|\boldsymbol{\varphi}^N - \boldsymbol{\theta}^N\|_0 = d_1}} p(\boldsymbol{\varphi}^N) f(d_1, 0; \gamma, Q, M) \left( \sum_{\mathbf{v}^M \in \mathbb{F}_Q^M} p(\mathbf{v}^M) \right) + 2\varepsilon \\ &\stackrel{(c)}{\leq} \sum_{d_1=1}^{\lfloor \alpha N \rfloor} \sum_{\substack{(\boldsymbol{\theta}^N, \mathbf{u}^M) \in \mathcal{U} \\ \boldsymbol{\varphi}^N \in \mathbb{F}_Q^N: \|\boldsymbol{\varphi}^N - \boldsymbol{\theta}^N\|_0 = d_1}} p(\boldsymbol{\varphi}^N) f(1, 0; \gamma, Q, M) \\ &\quad + \sum_{d_1=\lceil \alpha N \rceil}^N \sum_{\substack{(\boldsymbol{\theta}^N, \mathbf{u}^M) \in \mathcal{U} \\ \boldsymbol{\varphi}^N \in \mathbb{F}_Q^N: \|\boldsymbol{\varphi}^N - \boldsymbol{\theta}^N\|_0 = d_1}} p(\boldsymbol{\varphi}^N) f(\lceil \alpha N \rceil, 0; \gamma, Q, M) + 2\varepsilon, \end{aligned} \quad (59)$$

where (a) is by the classification of  $\boldsymbol{\varphi}^N$  and  $\mathbf{v}^M$  according to the  $\ell_0$  norm of their difference with  $\boldsymbol{\theta}^N$  and  $\mathbf{u}^M$  respectively and (b) is obtained using the bound (56) and

using  $\sum_{\mathbf{v}^M \in \mathbb{F}_Q^M} p(\mathbf{v}^M) = 1$ . The splitting in (c) permits  $f(d_1, 0; \gamma, Q, M)$  to be bounded in different cases, this idea comes from [27] and is also meaningful here. The parameter  $\alpha$  is a positive real number with  $0 < \alpha < 0.5$ . The way to choose  $\alpha$  is discussed in Section IV-B. The two terms in (59), denoted by  $P_{\mathcal{U}_1}(\alpha)$  and  $P_{\mathcal{U}_2}(\alpha)$ , need to be considered separately. For the first term  $P_{\mathcal{U}_1}(\alpha)$ , we have

$$\begin{aligned}
P_{\mathcal{U}_1}(\alpha) &= f(1, 0; \gamma, Q, M) \sum_{d_1=1}^{\lfloor \alpha N \rfloor} \sum_{\substack{(\boldsymbol{\theta}^N, \mathbf{u}^M) \in \mathcal{U} \\ \boldsymbol{\varphi}^N \in \mathbb{F}_Q^N : \|\boldsymbol{\varphi}^N - \boldsymbol{\theta}^N\|_0 = d_1}} p(\boldsymbol{\varphi}^N) \\
&\stackrel{(a)}{=} (1 - \gamma)^M \sum_{\mathbf{u}^M \in \mathcal{A}_{[u]\varepsilon}^M} \sum_{d_1=1}^{\lfloor \alpha N \rfloor} \sum_{\boldsymbol{\varphi}^N \in \mathbb{F}_Q^N} p(\boldsymbol{\varphi}^N) \sum_{\substack{\boldsymbol{\theta}^N \in \mathcal{A}_{[\theta]\varepsilon}^N : \\ \|\boldsymbol{\theta}^N - \boldsymbol{\varphi}^N\|_0 = d_1}} 1 \\
&\stackrel{(b)}{\leq} (1 - \gamma)^M \sum_{\mathbf{u}^M \in \mathcal{A}_{[u]\varepsilon}^M} \sum_{d_1=1}^{\lfloor \alpha N \rfloor} \sum_{\boldsymbol{\varphi}^N \in \mathbb{F}_Q^N} p(\boldsymbol{\varphi}^N) |\{\boldsymbol{\theta}^N \in \mathbb{F}_Q^N : \|\boldsymbol{\theta}^N - \boldsymbol{\varphi}^N\|_0 = d_1\}| \\
&\stackrel{(c)}{\leq} (1 - \gamma)^M \sum_{\mathbf{u}^M \in \mathcal{A}_{[u]\varepsilon}^M} \sum_{d_1=1}^{\lfloor \alpha N \rfloor} 2^{N H_2(\frac{d_1}{N})} (Q - 1)^{d_1} \\
&\stackrel{(d)}{\leq} (1 - \gamma)^M \cdot |\mathcal{A}_{[u]\varepsilon}^M| \cdot \alpha N \cdot 2^{N H_2(\alpha)} (Q - 1)^{\alpha N} \\
&\stackrel{(e)}{\leq} 2^{-N(-\frac{M}{N}(H(p_u) + \log(1 - \gamma) + \varepsilon) - H_2(\alpha) - \alpha \log(Q - 1) - \frac{\log(\alpha N)}{N})} = P_1(\alpha) \tag{60}
\end{aligned}$$

where (a) is by changing the order of summation and (b) is obtained considering all  $\boldsymbol{\theta}^N \in \mathbb{F}_Q^N$  and not only typical sequences. The bound (c) is obtained noticing that

$$\begin{aligned}
|\{\boldsymbol{\theta}^N \in \mathbb{F}_Q^N : \|\boldsymbol{\theta}^N - \boldsymbol{\varphi}^N\|_0 = d_1\}| &= \binom{N}{d_1} (Q - 1)^{d_1} \\
&\leq 2^{N H_2(\frac{d_1}{N})} (Q - 1)^{d_1}, \tag{61}
\end{aligned}$$

where  $H_2(p)$  denotes the entropy of a Bernoulli- $p$  source and  $\sum_{\boldsymbol{\varphi}^N \in \mathbb{F}_Q^N} p(\boldsymbol{\varphi}^N) = 1$ ; (d) is because of the monotonicity of the function  $H_2(\frac{d_1}{N})$ , which is increasing in  $d_1$  as  $d_1 \leq \lfloor \alpha N \rfloor < N/2$ ; (e) comes from [23, Theorem 3.1.2], the upper bound of the size

of  $\mathcal{A}_{[u]\varepsilon}^M$ , i.e.,

$$|\mathcal{A}_{[u]\varepsilon}^M| \leq 2^{M(H(p_u)+\varepsilon)}, \quad (62)$$

for  $M > M_\varepsilon$ . Similarly, for  $N > N_\varepsilon$ , one has

$$|\mathcal{A}_{[\Theta]\varepsilon}^N| \leq 2^{N(\mathcal{H}(\Theta)+\varepsilon)}. \quad (63)$$

Now we turn to  $P_{\mathcal{U}_2}(\alpha)$ ,

$$\begin{aligned} P_{\mathcal{U}_2}(\alpha) &= \sum_{d_1=\lceil \alpha N \rceil}^N \sum_{\substack{(\boldsymbol{\theta}^N, \mathbf{u}^M) \in \mathcal{U} \\ \boldsymbol{\varphi}^N \in \mathbb{F}_Q^N: \|\boldsymbol{\varphi}^N - \boldsymbol{\theta}^N\|_0 = d_1}} p(\boldsymbol{\varphi}^N) f(\lceil \alpha N \rceil, 0; \gamma, Q, M) \\ &\stackrel{(a)}{\leq} \sum_{(\boldsymbol{\theta}^N, \mathbf{u}^M) \in \mathcal{U}} \sum_{\boldsymbol{\varphi}^N \in \mathbb{F}_Q^N} p(\boldsymbol{\varphi}^N) f(\lceil \alpha N \rceil, 0; \gamma, Q, M) \\ &= |\mathcal{A}_{[\Theta]\varepsilon}^N| \cdot |\mathcal{A}_{[u]\varepsilon}^M| \cdot \left( Q^{-1} + \left( 1 - \frac{\gamma}{1 - Q^{-1}} \right)^{\lceil \alpha N \rceil} (1 - Q^{-1}) \right)^M \\ &\stackrel{(b)}{\leq} 2^{-N \left( -\mathcal{H}(\Theta) - \frac{M}{N} \left( H(p_u) + \log \left( Q^{-1} + \left( 1 - \frac{\gamma}{1 - Q^{-1}} \right)^{\lceil \alpha N \rceil} (1 - Q^{-1}) \right) + \varepsilon \right) - \varepsilon \right)} = P_2(\alpha) \end{aligned} \quad (64)$$

where (a) is by ignoring the constraint that  $\|\boldsymbol{\varphi}^N - \boldsymbol{\theta}^N\|_0 = d_1$ , and (b) is by the upper bounds of  $|\mathcal{A}_{[\Theta]\varepsilon}^N|$  and  $|\mathcal{A}_{[u]\varepsilon}^M|$ , as before. Equations (59), (60), and (64) complete the proof.  $\blacksquare$

### B. Sufficient Condition

In this section, sufficient conditions for the WN and NC cases are derived to get a vanishing upper bound of error probability.

**Proposition 3** (Sufficient condition, WN and NC cases). *Assume the absence of sensing noise and consider a sensing matrix with sparsity factor  $\gamma$ . For some  $\delta \in \mathbb{R}^+$  (which may be taken arbitrary close to zero), there exists small positive real numbers  $\varepsilon$ ,  $\xi$ , and integers  $N_\delta$ ,  $M_\varepsilon$  such that  $\forall N > N_\delta$  and  $M > M_\varepsilon$ , if the following conditions hold*

- the communication noise is not uniformly distributed, i.e.,

$$H(p_u) < \log Q - \xi, \quad (65)$$

- the sparsity factor is lower bounded

$$\gamma > 1 - 2^{-H(p_u) - \varepsilon}, \quad (66)$$

- the compression ratio  $M/N$  satisfies

$$\frac{M}{N} > \frac{\mathcal{H}(\Theta) + \varepsilon}{\log Q - H(p_u) - \xi}, \quad (67)$$

then one has  $P_e \leq \delta$  using MAP decoding. As  $N \rightarrow \infty$  and  $M \rightarrow \infty$ ,  $\varepsilon$  and  $\xi$  can be chosen arbitrary close to zero.

*Proof:* Both  $P_1(\alpha)$  and  $P_2(\alpha)$  need to be vanishing for increasing  $N$  and  $M$ . The exponent of each term is considered respectively. Define, from (60),

$$E_1^{\text{NC}} = -\frac{M}{N} (H(p_u) + \log(1 - \gamma) + \varepsilon) - H_2(\alpha) - \alpha \log(Q - 1) - \frac{\log(\alpha N)}{N}. \quad (68)$$

Then  $\lim_{N \rightarrow \infty} 2^{-NE_1^{\text{NC}}} = 0$  if  $E_1^{\text{NC}} > 0$ . Thus, if  $E_1^{\text{NC}} > 0$ , for any  $\tau_1 \in \mathbb{R}^+$  arbitrarily small,  $\exists N_{\tau_1}$  such that  $\forall N > N_{\tau_1}$ , one has  $P_1(\alpha) < \tau_1$ .

Notice that if  $H(p_u) + \log(1 - \gamma) + \varepsilon \geq 0$ ,  $E_1^{\text{NC}}$  is negative, thus one should first have

$$H(p_u) + \log(1 - \gamma) + \varepsilon < 0, \quad (69)$$

leading to (66). With this condition,  $E_1^{\text{NC}} > 0$  leads to

$$\frac{M}{N} > \frac{H_2(\alpha) + \alpha \log(Q - 1) + \frac{\log(\alpha N)}{N}}{\log \frac{1}{1 - \gamma} - H(p_u) - \varepsilon}. \quad (70)$$

Similarly, define from (64)

$$E_2^{\text{NC}} = -\mathcal{H}(\Theta) - \frac{M}{N} \left( H(p_u) + \log \left( Q^{-1} + \left( 1 - \frac{\gamma}{1 - Q^{-1}} \right)^{\lceil \alpha N \rceil} (1 - Q^{-1}) \right) + \varepsilon \right) - \varepsilon. \quad (71)$$

Again, if  $E_2^{\text{NC}} > 0$ , for any  $\tau_2 \in \mathbb{R}^+$  arbitrarily small,  $\exists N_{\tau_2} \in \mathbb{N}^+$  such that  $\forall N > N_{\tau_2}$ , one has  $P_2(\alpha) < \tau_2$ . Since  $0 < \gamma \leq 1 - Q^{-1}$ , one gets  $0 \leq 1 - \frac{\gamma}{1-Q^{-1}} < 1$  and

$$\lim_{N \rightarrow \infty} \left(1 - \frac{\gamma}{1-Q^{-1}}\right)^{[\alpha N]} = 0. \quad (72)$$

Thus for  $\sigma \in \mathbb{R}^+$  arbitrarily small, there exists an  $N_\sigma$  such that for  $\forall N > N_\sigma$ ,

$$\left(1 - \frac{\gamma}{1-Q^{-1}}\right)^{[\alpha N]} (1 - Q^{-1}) < \sigma Q^{-1}. \quad (73)$$

Hence  $E_2^{\text{NC}}$  in (71) can be lower bounded by

$$E_2^{\text{NC}} > -\mathcal{H}(\Theta) - \frac{M}{N} (H(p_u) + \log(Q^{-1} + \sigma Q^{-1}) + \varepsilon) - \varepsilon, \quad (74)$$

for  $N > N_\sigma$ . If this lower bound is positive, then  $E_2^{\text{NC}}$  is positive. Again, if  $H(p_u) - \log Q + \log(1 + \sigma) + \varepsilon \leq 0$ , one obtains a negative lower bound for  $E_2^{\text{NC}}$  from (74). Thus, one deduces (65) in Proposition 3, with

$$\xi = \log(1 + \sigma) + \varepsilon. \quad (75)$$

From (65), to get a positive lower bound for (74), one should have

$$\frac{M}{N} > \frac{\mathcal{H}(\Theta) + \varepsilon}{\log Q - H(p_u) - \log(1 + \sigma) - \varepsilon}. \quad (76)$$

From (76) and (75), with  $\xi \rightarrow 0$  as  $N \rightarrow \infty$ , one gets (67) in Proposition 3.

From (70) and (76), one obtains

$$\frac{M}{N} > \max \left\{ \frac{H_2(\alpha) + \alpha \log(Q - 1) + \frac{\log(\alpha N)}{N}}{\log \frac{1}{1-\gamma} - H(p_u) - \varepsilon}, \frac{\mathcal{H}(\Theta) + \varepsilon}{\log Q - H(p_u) - \xi} \right\}. \quad (77)$$

The value of  $\alpha$  should be chosen such that the lower bound (77) on  $M/N$  is minimum. One may compare (77) with the necessary condition (14). The second term of (77) is similar to (14), since both  $\xi$  and  $\varepsilon$  can be made arbitrarily close to 0 as  $N \rightarrow \infty$ . The best value for  $\alpha$  has thus to be such that

$$\frac{H_2(\alpha) + \alpha \log(Q - 1) + \frac{\log(\alpha N)}{N}}{\log \frac{1}{1-\gamma} - H(p_u) - \varepsilon} \leq \frac{\mathcal{H}(\Theta) + \varepsilon}{\log Q - H(p_u) - \xi}. \quad (78)$$



The function  $H_2(\alpha) + \alpha \log(Q - 1)$  is increasing when  $\alpha \in ]0, 0.5[$  and tends to 0 as  $\alpha \rightarrow 0$ . The term  $\log(\alpha N)/N$  is also negligible for  $N$  large. Thus, there always exists some  $\alpha$  satisfying (78). Since the speed of convergence of  $\xi$  is affected by  $\alpha$ , we choose the largest  $\alpha$  that satisfies (78). Finally, the sufficient condition (67) is obtained for  $M/N$ .

From (31), one may conclude that

$$P_e \leq \tau_1 + \tau_2 + 2\varepsilon. \quad (79)$$

To ensure  $P_e < \delta$ , we should choose  $\tau_1$ ,  $\tau_2$ , and  $\varepsilon$  to satisfy  $\tau_1 + \tau_2 + 2\varepsilon < \delta$ . Then a proper value of  $\sigma$ , which depends on  $\tau_2$  and  $\varepsilon$ , can be chosen. At last,  $\xi$  is obtained from (75). With these well determined parameters, if all the three conditions in Proposition 3 hold, there exists integers  $N_\varepsilon$ ,  $N_{\tau_1}$ ,  $N_{\tau_2}$ , and  $N_\sigma$ , such that for any

$$N > N_\delta = \max \{N_\varepsilon, N_{\tau_1}, N_{\tau_2} N_\sigma\}, \quad (80)$$

and  $M > M_\varepsilon$ , one has  $P_e < \delta$ . ■

### C. Discussion and Numerical Results

In [18, Eq. (24)], considering a sparse and iid source, a uniformly distributed random matrix  $\mathbf{A}$ , and the minimum empirical entropy decoder, the following error exponent in the case NC is obtained

$$E_0^{\text{NC}} = \min_{p,q} D(p \parallel p_\Theta) + \frac{M}{N} D(q \parallel p_u) + \left| \frac{M}{N} \log Q - H(p) - \frac{M}{N} H(q) \right|^+, \quad (81)$$

where  $D(\cdot \parallel \cdot)$  denotes the relative entropy between two distributions and  $|\cdot|^+ = \max\{0, \cdot\}$ .

In parallel, [12] proposed an approach to prove that the upper bound for the probability of decoding error  $P_e$  under minimum empirical entropy decoding is equal to that of the maximum  $Q$ -probability decoder. As discussed in Section IV-A, in the WN and NC cases, the MAP decoder in the considered context is equivalent to the maximum

$Q$ -probability decoder. As a consequence, (81) is also the error exponent of the MAP decoder in the considered context. A proof for (81) using the method of types need to do some assumptions on the topology of the considered sensor network to specify the type of  $\theta^N$ . For correlated sources, one can extend (81) considering Markov model, and use higher-order types, leading to cumbersome derivations.

From (81), provided that  $E_0^{\text{NC}} > 0$ ,  $P_e$  tends to 0 as  $N$  increases.  $E_0^{\text{NC}}$  cannot be negative and  $E_0^{\text{NC}} = 0$  if and only if

$$\begin{cases} D(p \parallel p_\Theta) = 0, \\ D(q \parallel p_u) = 0, \\ \frac{M}{N} \log Q - H(p) - \frac{M}{N} H(q) \leq 0. \end{cases} \quad (82)$$

Thus, (82) implies that  $\frac{M}{N} \log Q - H(p_\Theta) - \frac{M}{N} H(p_u) \leq 0$ . Thus, a necessary and sufficient condition to have  $E_0^{\text{NC}} > 0$  is  $\frac{M}{N} \log Q - H(p_\Theta) - \frac{M}{N} H(p_u) > 0$ , which is the same as (67) with  $\gamma = 1 - Q^{-1}$  (corresponding to  $\mathbf{A}$  uniformly distributed). The proof using weak typicality leads to the same results (in terms of sufficient condition for having asymptotically vanishing  $P_e$ ) as the technique in [18].

In the noiseless case, since  $\gamma$  can be chosen arbitrarily small, the necessary condition in Proposition 1 and the sufficient condition in Proposition 3 asymptotically coincide. This confirms the numerical results obtained in [19]. In the NC case, the difference between the two conditions comes from the constraint linking  $\gamma$  and the entropy of the communication noise. In Section III, the structure of  $\mathbf{A}$  was not considered and no condition on  $\gamma$  has been obtained. The lower bound on  $\gamma$  implies that  $\mathbf{A}$  should be dense enough to fight against the noise. Since the communication noise is iid, for a given probability of having one entry of  $\mathbf{u}^M$  non-zero, *i.e.*,  $\Pr(u \neq 0)$ , the entropy  $H(p_u)$  is maximized when  $p_u(q) = \Pr(u \neq 0) / (Q - 1)$  for any  $q \in \mathbb{F}_Q \setminus \{0\}$ . This corresponds to the worst noise in terms of compression efficiency.

Figure 2 represents the lower bound of  $\gamma$  as a function of  $\Pr(u \neq 0)$ , ranging from

$10^{-5}$  to  $10^{-1}$ , for different value of  $Q$ . There is almost no requirement on  $\gamma$  when  $\Pr(u \neq 0) \leq 5 \times 10^{-4}$ . For a given noise level, a larger size of the finite field needs a denser sensing matrix. Figure 3 shows the influence of the communication noise on the optimum compression ratio. The lower bound of  $M/N$  is represented as a function of  $\mathcal{H}(\Theta) / \log Q$ , for different values of  $Q$  and for different values of  $\Pr(u \neq 0)$ .

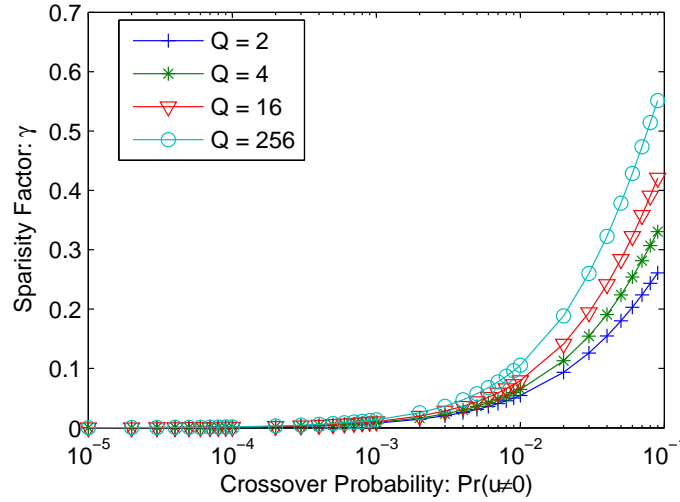


Figure 2. Lower bound of  $\gamma$  to achieve the optimum compression ratio for  $N \rightarrow \infty$ , according to (66)

## V. SUFFICIENT CONDITION IN PRESENCE OF SENSING NOISE

This section performs an achievability study in presence of sensing noise by considering the conditional pmf  $p_{\mathbf{x}|\Theta}$ . The communication noise  $\mathbf{u}^M$  is first neglected to simplify the problem (NS case). The extension to the NCS case is easily obtained from the NS case. Assume that  $\boldsymbol{\theta}^N$  is the true state vector and that  $\mathbf{x}^N$  represents the measurements of the sensors. The sink receives  $\mathbf{y}^M = \mathbf{A}\mathbf{x}^N$ . The *a posteriori* pmf (7) can be written

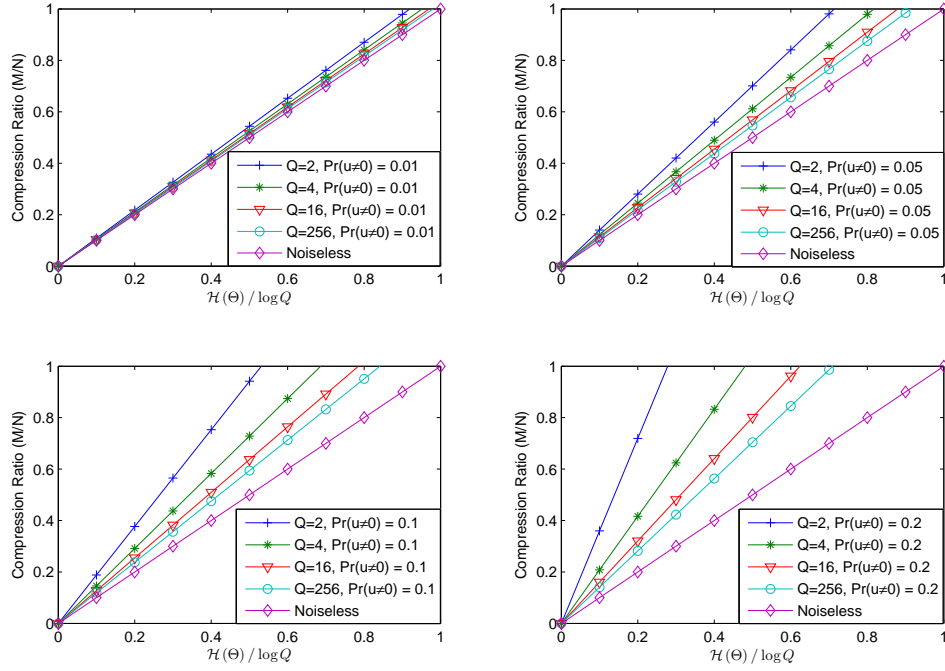


Figure 3. Optimum asymptotic achievable compression ratio in function of  $\mathcal{H}(\Theta) / \log Q$ , according to (67), for a crossover probability equal to 0.01, 0.05, 0.1, and 0.2 respectively, and without noise

as

$$p(\boldsymbol{\theta}^N | \mathbf{y}^M, \mathbf{A}) \propto \sum_{\mathbf{z}^N \in \mathbb{F}_Q^N} p(\boldsymbol{\theta}^N) p(\mathbf{z}^N | \boldsymbol{\theta}^N) 1_{\mathbf{y}^M = \mathbf{A}\mathbf{z}^N}. \quad (83)$$

In the case of MAP estimation, an error occurs if there exists a vector  $\boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\}$  such that

$$\sum_{\mathbf{z}^N \in \mathbb{F}_Q^N} p(\boldsymbol{\theta}^N, \mathbf{z}^N) 1_{\mathbf{y}^M = \mathbf{A}\mathbf{z}^N} \leq \sum_{\mathbf{z}^N \in \mathbb{F}_Q^N} p(\boldsymbol{\varphi}^N, \mathbf{z}^N) 1_{\mathbf{y}^M = \mathbf{A}\mathbf{z}^N}. \quad (84)$$

$\boldsymbol{\theta}^N$  and  $\mathbf{x}^N$  are considered as fixed, but unknown. The decoder has knowledge of  $\mathbf{A}$  and

$\mathbf{y}^M = \mathbf{A}\mathbf{x}^N$ , thus an alternative way to express (84) is

$$\sum_{\mathbf{z}^N \in \mathbb{F}_Q^N} p(\boldsymbol{\theta}^N, \mathbf{z}^N) 1_{\mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{z}^N} \leq \sum_{\mathbf{z}^N \in \mathbb{F}_Q^N} p(\boldsymbol{\varphi}^N, \mathbf{z}^N) 1_{\mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{z}^N}. \quad (85)$$

#### A. Achievability Study

We begin with the extension of the basic weakly typical set as introduced in Section IV-A. For any  $\varepsilon > 0$  and  $N \in \mathbb{N}^+$ , based on  $\mathcal{A}_{[\Theta]\varepsilon}^N$  for  $\boldsymbol{\theta}^N$ , one defines the weakly conditional typical set  $\mathcal{A}_{[\mathbf{x}|\Theta]\varepsilon}^N(\boldsymbol{\theta}^N)$  for  $\mathbf{x}^N$ , which is conditionally distributed with respect to  $p_{\mathbf{x}|\Theta}$ , with  $\boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]\varepsilon}^N$ ,

$$\mathcal{A}_{[\mathbf{x}|\Theta]\varepsilon}^N(\boldsymbol{\theta}^N) = \left\{ \mathbf{x}^N \in \mathbb{F}_Q^N \text{ such that } \left| -\frac{1}{N} \log p(\mathbf{x}^N | \boldsymbol{\theta}^N) - \mathcal{H}(\mathbf{x} | \Theta) \right| \leq \varepsilon \right\}. \quad (86)$$

Since  $\mathcal{H}(\Theta, \mathbf{x}) = \mathcal{H}(\Theta) + \mathcal{H}(\mathbf{x} | \Theta)$ , if  $\boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]\varepsilon}^N$  and  $\mathbf{x}^N \in \mathcal{A}_{[\mathbf{x}|\Theta]\varepsilon}^N(\boldsymbol{\theta}^N)$ , then  $(\boldsymbol{\theta}^N, \mathbf{x}^N) \in \mathcal{A}_{[\Theta, \mathbf{x}]2\varepsilon}^N$  by consistency, where  $\mathcal{A}_{[\Theta, \mathbf{x}]2\varepsilon}^N$  denotes the weakly joint typical set, *i.e.*, the set of pairs  $(\boldsymbol{\theta}^N, \mathbf{x}^N) \in \mathbb{F}_Q^N \times \mathbb{F}_Q^N$  such that

$$\left| -\frac{1}{N} \log p(\boldsymbol{\theta}^N, \mathbf{x}^N) - \mathcal{H}(\Theta, \mathbf{x}) \right| \leq 2\varepsilon. \quad (87)$$

For any  $\varepsilon > 0$  there exist an  $N_\varepsilon$  such that for all  $N \geq N_\varepsilon$  and for any  $\boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]\varepsilon}^N$ , one has  $\Pr \left\{ \mathbf{x}^N \in \mathcal{A}_{[\mathbf{x}|\Theta]\varepsilon}^N(\boldsymbol{\theta}^N) \right\} \geq 1 - \varepsilon$  and  $\Pr \left\{ (\boldsymbol{\theta}^N, \mathbf{x}^N) \in \mathcal{A}_{[\Theta, \mathbf{x}]2\varepsilon}^N \right\} \geq 1 - 2\varepsilon$ . The cardinality of the set  $\mathcal{A}_{[\Theta, \mathbf{x}]2\varepsilon}^N$  satisfies

$$|\mathcal{A}_{[\Theta, \mathbf{x}]2\varepsilon}^N| \leq 2^{N(\mathcal{H}(\Theta, \mathbf{x}) + 2\varepsilon)}. \quad (88)$$

One may have  $\varepsilon$  arbitrary close to zero as  $N \rightarrow \infty$ .

Considering  $\mathcal{A}_{[\Theta, \mathbf{x}]2\varepsilon}^N$ , the estimation error probability is bounded by

$$\begin{aligned} P_e &\leq \sum_{(\boldsymbol{\theta}^N, \mathbf{x}^N) \in \mathcal{A}_{[\Theta, \mathbf{x}]2\varepsilon}^N} p(\boldsymbol{\theta}^N, \mathbf{x}^N) \Pr \{ \text{error} | \boldsymbol{\theta}^N, \mathbf{x}^N \} + \sum_{(\boldsymbol{\theta}^N, \mathbf{x}^N) \notin \mathcal{A}_{[\Theta, \mathbf{x}]2\varepsilon}^N} p(\boldsymbol{\theta}^N, \mathbf{x}^N) \\ &\leq \sum_{(\boldsymbol{\theta}^N, \mathbf{x}^N) \in \mathcal{A}_{[\Theta, \mathbf{x}]2\varepsilon}^N} p(\boldsymbol{\theta}^N, \mathbf{x}^N) \cdot \Pr \{ \text{error} | \boldsymbol{\theta}^N, \mathbf{x}^N \} + 2\varepsilon, \end{aligned} \quad (89)$$

Errors appear mainly due to a bad sensing matrix. Averaging over all  $\mathbf{A} \in \mathbb{F}_Q^{M \times N}$ , (89) becomes

$$P_e \leq \sum_{\mathbf{A} \in \mathbb{F}_Q^{M \times N}} p(\mathbf{A}) \sum_{\boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]_\varepsilon}^N} \sum_{\mathbf{x}^N \in \mathcal{A}_{[\mathbf{x}|\Theta]_\varepsilon}^N(\boldsymbol{\theta}^N)} p(\boldsymbol{\theta}^N, \mathbf{x}^N) \Pr\{\text{error} \mid \boldsymbol{\theta}^N, \mathbf{x}^N, \mathbf{A}\} + 2\varepsilon, \quad (90)$$

where  $p(\mathbf{A}) = \Pr\{\mathbf{A} = \mathbf{A}\}$ .  $\Pr\{\text{error} \mid \boldsymbol{\theta}^N, \mathbf{x}^N, \mathbf{A}\}$  can be written as

$$\Pr\{\text{error} \mid \boldsymbol{\theta}^N, \mathbf{x}^N, \mathbf{A}\} = \begin{cases} 1 & \text{if } \exists \boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\} \text{ s.t. (85) holds,} \\ 0 & \text{if } \forall \boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\}, \text{ (85) does not hold.} \end{cases} \quad (91)$$

Using again the idea of Lemma 1, the conditional error probability is bounded by

$$\Pr\{\text{error} \mid \boldsymbol{\theta}^N, \mathbf{x}^N, \mathbf{A}\} \leq \sum_{\boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\}} \frac{\sum_{\mathbf{z}_1^N \in \mathbb{F}_Q^N} p(\boldsymbol{\varphi}^N, \mathbf{z}_1^N) 1_{\mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{z}_1^N}}{\sum_{\mathbf{z}_2^N \in \mathbb{F}_Q^N} p(\boldsymbol{\theta}^N, \mathbf{z}_2^N) 1_{\mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{z}_2^N}}. \quad (92)$$

From (90) and (92), one gets

$$P_e \leq \sum_{\substack{\mathbf{A} \in \mathbb{F}_Q^{M \times N} \\ \boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]_\varepsilon}^N}} p(\mathbf{A}) \sum_{\mathbf{x}^N \in \mathcal{A}_{[\mathbf{x}|\Theta]_\varepsilon}^N(\boldsymbol{\theta}^N)} p(\boldsymbol{\theta}^N, \mathbf{x}^N) \sum_{\boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\}} \frac{\sum_{\mathbf{z}_1^N \in \mathbb{F}_Q^N} p(\boldsymbol{\varphi}^N, \mathbf{z}_1^N) 1_{\mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{z}_1^N}}{\sum_{\mathbf{z}_2^N \in \mathbb{F}_Q^N} p(\boldsymbol{\theta}^N, \mathbf{z}_2^N) 1_{\mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{z}_2^N}} + 2\varepsilon. \quad (93)$$

Now, for some  $\boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]_\varepsilon}^N$ , consider the direct image by  $\mathbf{A}$  of the conditional typical set  $\mathcal{A}_{[\mathbf{x}|\Theta]_\varepsilon}^N(\boldsymbol{\theta}^N)$

$$\mathcal{Y}_\varepsilon(\mathbf{A}, \boldsymbol{\theta}^N) = \{\mathbf{y}^M = \mathbf{A}\mathbf{x}^N, \text{ for all } \mathbf{x}^N \in \mathcal{A}_{[\mathbf{x}|\Theta]_\varepsilon}^N(\boldsymbol{\theta}^N)\}. \quad (94)$$

**Lemma 3.** For any arbitrary real-valued function  $h(\mathbf{x}^N)$  with  $\mathbf{x}^N \in \mathbb{F}_Q^N$ , one has

$$\sum_{\mathbf{x}^N \in \mathcal{A}_{[\mathbf{x}|\Theta]_\varepsilon}^N(\boldsymbol{\theta}^N)} h(\mathbf{x}^N) = \sum_{\mathbf{y}^M \in \mathcal{Y}_\varepsilon(\mathbf{A}, \boldsymbol{\theta}^N)} \sum_{\mathbf{x}^N \in \mathcal{A}_{[\mathbf{x}|\Theta]_\varepsilon}^N(\boldsymbol{\theta}^N)} h(\mathbf{x}^N) 1_{\mathbf{y}^M = \mathbf{A}\mathbf{x}^N}. \quad (95)$$

*Proof:* For a given  $\mathbf{y}^M \in \mathcal{Y}_\varepsilon(\mathbf{A}, \boldsymbol{\theta}^N)$ , consider the set

$$\mathcal{X}_\varepsilon(\mathbf{y}^M, \mathbf{A}, \boldsymbol{\theta}^N) = \{\mathbf{x}^N \in \mathcal{A}_{[\mathbf{x}|\Theta]_\varepsilon}^N(\boldsymbol{\theta}^N) \text{ such that } \mathbf{y}^M = \mathbf{A}\mathbf{x}^N\}. \quad (96)$$

Then one has

$$\mathcal{A}_{[x|\Theta]^\varepsilon}^N(\theta^N) = \bigcup_{\mathbf{y}^M \in \mathcal{Y}_\varepsilon(\mathbf{A}, \theta^N)} \mathcal{X}_\varepsilon(\mathbf{y}^M, \mathbf{A}, \theta^N), \quad (97)$$

with  $\mathcal{X}_\varepsilon(\mathbf{y}_i^M, \mathbf{A}, \theta^N) \cap \mathcal{X}_\varepsilon(\mathbf{y}_j^M, \mathbf{A}, \theta^N) = \emptyset$  for any  $\mathbf{y}_i^M \neq \mathbf{y}_j^M$ , since the multiplication by  $\mathbf{A}$  is a surjection from  $\mathcal{A}_{[x|\Theta]^\varepsilon}^N(\theta^N)$  to  $\mathcal{Y}_\varepsilon(\mathbf{A}, \theta^N)$ . So any sum over  $\mathbf{x}^N \in \mathcal{A}_{[x|\Theta]^\varepsilon}^N(\theta^N)$  can be decomposed as

$$\begin{aligned} \sum_{\mathbf{x}^N \in \mathcal{A}_{[x|\Theta]^\varepsilon}^N(\theta^N)} h(\mathbf{x}^N) &= \sum_{\mathbf{y}^M \in \mathcal{Y}_\varepsilon(\mathbf{A}, \theta^N)} \sum_{\mathbf{x}^N \in \mathcal{X}_\varepsilon(\mathbf{y}^M, \mathbf{A}, \theta^N)} h(\mathbf{x}^N) \\ &= \sum_{\mathbf{y}^M \in \mathcal{Y}_\varepsilon(\mathbf{A}, \theta^N)} \sum_{\mathbf{x}^N \in \mathcal{A}_{[x|\Theta]^\varepsilon}^N(\theta^N)} h(\mathbf{x}^N) 1_{\mathbf{y}^M = \mathbf{A}\mathbf{x}^N}. \end{aligned} \quad (98)$$

■

Applying (95) to (93), one obtains

$$\begin{aligned} P_e &\leq \sum_{\substack{\mathbf{A} \in \mathbb{F}_Q^{M \times N} \\ \theta^N \in \mathcal{A}_{[\Theta]^\varepsilon}^N}} p(\mathbf{A}) \sum_{\mathbf{y}^M \in \mathcal{Y}_\varepsilon(\mathbf{A}, \theta^N)} \sum_{\mathbf{x}^N \in \mathcal{A}_{[x|\Theta]^\varepsilon}^N(\theta^N)} p(\theta^N, \mathbf{x}^N) 1_{\mathbf{y}^M = \mathbf{A}\mathbf{x}^N} \\ &\quad \cdot \left( \sum_{\varphi^N \in \mathbb{F}_Q^N \setminus \{\theta^N\}} \frac{\sum_{\mathbf{z}_1^N \in \mathbb{F}_Q^N} p(\varphi^N, \mathbf{z}_1^N) 1_{\mathbf{y}^M = \mathbf{A}\mathbf{z}_1^N}}{\sum_{\mathbf{z}_2^N \in \mathbb{F}_Q^N} p(\theta^N, \mathbf{z}_2^N) 1_{\mathbf{y}^M = \mathbf{A}\mathbf{z}_2^N}} \right) + 2\varepsilon \\ &= \sum_{\substack{\mathbf{A} \in \mathbb{F}_Q^{M \times N} \\ \theta^N \in \mathcal{A}_{[\Theta]^\varepsilon}^N}} p(\mathbf{A}) \sum_{\mathbf{y}^M \in \mathcal{Y}_\varepsilon(\mathbf{A}, \theta^N)} \left( \sum_{\varphi^N \in \mathbb{F}_Q^N \setminus \{\theta^N\}} \sum_{\mathbf{z}_1^N \in \mathbb{F}_Q^N} p(\varphi^N, \mathbf{z}_1^N) \cdot 1_{\mathbf{y}^M = \mathbf{A}\mathbf{z}_1^N} \right) \\ &\quad \cdot \left( \frac{\sum_{\mathbf{x}^N \in \mathcal{A}_{[x|\Theta]^\varepsilon}^N(\theta^N)} p(\theta^N, \mathbf{x}^N) 1_{\mathbf{y}^M = \mathbf{A}\mathbf{x}^N}}{\sum_{\mathbf{z}_2^N \in \mathbb{F}_Q^N} p(\theta^N, \mathbf{z}_2^N) 1_{\mathbf{y}^M = \mathbf{A}\mathbf{z}_2^N}} \right) + 2\varepsilon \\ &\leq \sum_{\substack{\mathbf{A} \in \mathbb{F}_Q^{M \times N} \\ \theta^N \in \mathcal{A}_{[\Theta]^\varepsilon}^N}} p(\mathbf{A}) \sum_{\mathbf{y}^M \in \mathcal{Y}_\varepsilon(\mathbf{A}, \theta^N)} \left( \sum_{\varphi^N \in \mathbb{F}_Q^N \setminus \{\theta^N\}} \sum_{\mathbf{z}_1^N \in \mathbb{F}_Q^N} p(\varphi^N, \mathbf{z}_1^N) 1_{\mathbf{y}^M = \mathbf{A}\mathbf{z}_1^N} \right) + 2\varepsilon, \end{aligned} \quad (99)$$

since we have

$$\frac{\sum_{\mathbf{x}^N \in \mathcal{A}_{[x|\Theta]\varepsilon}^N} p(\boldsymbol{\theta}^N) p(\mathbf{x}^N) 1_{\mathbf{y}^M = \mathbf{A}\mathbf{x}^N}}{\sum_{\mathbf{z}_2^N \in \mathbb{F}_Q^N} p(\boldsymbol{\theta}^N, \mathbf{z}_2^N) 1_{\mathbf{y}^M = \mathbf{A}\mathbf{z}_2^N}} \leq 1. \quad (100)$$

The bound (100) is tight because for  $N$  sufficiently large, the probability of the non-typical set vanishes. Recall that  $\mathbf{y}^M = \mathbf{A}\mathbf{x}^N$ , even though  $\mathbf{x}^N$  is not explicit in (99). As a vector  $\mathbf{y}^M$  may correspond to several  $\mathbf{x}^N$ s, (99) is further bounded by

$$\begin{aligned} P_e &\leq \sum_{\substack{\mathbf{A} \in \mathbb{F}_Q^{M \times N} \\ \boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]\varepsilon}^N}} p(\mathbf{A}) \sum_{\mathbf{x}^N \in \mathcal{A}_{[x|\Theta]\varepsilon}^N} \sum_{\substack{\boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\} \\ \mathbf{z}_1^N \in \mathbb{F}_Q^N}} p(\boldsymbol{\varphi}^N, \mathbf{z}_1^N) 1_{\mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{z}_1^N} + 2\varepsilon \\ &\leq \sum_{\substack{\boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]\varepsilon}^N \\ \boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\}}} \sum_{\substack{\mathbf{x}^N \in \mathcal{A}_{[x|\Theta]\varepsilon}^N(\boldsymbol{\theta}^N) \\ \mathbf{z}_1^N \in \mathbb{F}_Q^N}} p(\boldsymbol{\varphi}^N, \mathbf{z}_1^N) \sum_{\mathbf{A} \in \mathbb{F}_Q^{M \times N}} p(\mathbf{A}) 1_{\mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{z}_1^N} + 2\varepsilon. \end{aligned} \quad (101)$$

Since

$$\sum_{\mathbf{A} \in \mathbb{F}_Q^{M \times N}} p(\mathbf{A}) 1_{\mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{z}_1^N} = \Pr \{ \mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{z}_1^N \}, \quad (102)$$

one gets

$$P_e \leq \sum_{\substack{\boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]\varepsilon}^N \\ \boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\}}} \sum_{\substack{\mathbf{x}^N \in \mathcal{A}_{[x|\Theta]\varepsilon}^N(\boldsymbol{\theta}^N) \\ \mathbf{z}_1^N \in \mathbb{F}_Q^N}} p(\boldsymbol{\varphi}^N, \mathbf{z}_1^N) \Pr \{ \mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{z}_1^N \} + 2\varepsilon. \quad (103)$$

Suppose that  $\|\mathbf{x}^N - \mathbf{z}_1^N\|_0 = d$ . If  $d = 0$ ,  $\Pr \{ \mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{z}_1^N \}$  equals 1. Otherwise we can apply Lemma 2, without communication noise,  $\Pr \{ \mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{z}_1^N \} = f(d, 0; \gamma, Q, M)$ . Depending on  $d$  being zero or not,  $P_A$  is split as follows

$$P_e \leq P_{A_1} + P_{A_2} + 2\varepsilon, \quad (104)$$

where

$$P_{A_1} = \sum_{\substack{\boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]\varepsilon}^N \\ \boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\}}} \sum_{\mathbf{z}_1^N \in \mathcal{A}_{[x|\Theta]\varepsilon}^N(\boldsymbol{\theta}^N)} p(\boldsymbol{\varphi}^N, \mathbf{z}_1^N), \quad (105)$$



and

$$P_{\mathcal{A}_2} = \sum_{\substack{\boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]}^N \\ \boldsymbol{\varphi}^N \in \mathbb{F}_Q^N \setminus \{\boldsymbol{\theta}^N\}}} \sum_{\substack{\mathbf{x}^N \in \mathcal{A}_{[X|\Theta]}^N(\boldsymbol{\theta}^N) \\ \mathbf{z}_1^N \in \mathbb{F}_Q^N \setminus \{\mathbf{x}^N\}}} p(\boldsymbol{\varphi}^N, \mathbf{z}_1^N) \Pr\{\mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{z}_1^N\}. \quad (106)$$

**Lemma 4.** A sufficient condition for  $P_{\mathcal{A}_1} \leq 2\varepsilon$  is that, for any pair of vectors  $(\boldsymbol{\theta}^N, \boldsymbol{\varphi}^N) \in \mathcal{A}_{[\Theta]}^N \times \mathcal{A}_{[\Theta]}^N$  such that  $\boldsymbol{\theta}^N \neq \boldsymbol{\varphi}^N$ ,

$$\mathcal{A}_{[X|\Theta]}^N(\boldsymbol{\theta}^N) \cap \mathcal{A}_{[X|\Theta]}^N(\boldsymbol{\varphi}^N) = \emptyset. \quad (107)$$

*Proof:* Assume that (107) is satisfied. Changing the order of summation, (105) becomes

$$P_{\mathcal{A}_1} = \sum_{\boldsymbol{\varphi}^N \in \mathbb{F}_Q^N} p(\boldsymbol{\varphi}^N) \sum_{\substack{\boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]}^N \setminus \{\boldsymbol{\varphi}^N\} \\ \mathbf{z}_1^N \in \mathcal{A}_{[X|\Theta]}^N(\boldsymbol{\theta}^N)}} p(\mathbf{z}_1^N | \boldsymbol{\varphi}^N), \quad (108)$$

which can be further decomposed as  $P_{\mathcal{A}_1} = P_{\mathcal{A}_{11}} + P_{\mathcal{A}_{12}}$ , with

$$\begin{aligned} P_{\mathcal{A}_{11}} &= \sum_{\boldsymbol{\varphi}^N \in \mathcal{A}_{[\Theta]}^N} p(\boldsymbol{\varphi}^N) \sum_{\substack{\boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]}^N \setminus \{\boldsymbol{\varphi}^N\} \\ \mathbf{z}_1^N \in \mathcal{A}_{[X|\Theta]}^N(\boldsymbol{\theta}^N)}} p(\mathbf{z}_1^N | \boldsymbol{\varphi}^N) \\ &\stackrel{(a)}{\leq} \sum_{\boldsymbol{\varphi}^N \in \mathcal{A}_{[\Theta]}^N} p(\boldsymbol{\varphi}^N) \sum_{\mathbf{z}_1^N \in \mathbb{F}_Q^N \setminus \mathcal{A}_{[X|\Theta]}^N(\boldsymbol{\varphi}^N)} p(\mathbf{z}_1^N | \boldsymbol{\varphi}^N) \\ &\leq \sum_{\boldsymbol{\varphi}^N \in \mathcal{A}_{[\Theta]}^N} p(\boldsymbol{\varphi}^N) \varepsilon \leq \varepsilon, \end{aligned} \quad (109)$$

where (a) comes from the fact that if (107) is satisfied, one has

$$\bigcup_{\boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]}^N \setminus \{\boldsymbol{\varphi}^N\}} \mathcal{A}_{[X|\Theta]}^N(\boldsymbol{\theta}^N) \subseteq \mathbb{F}_Q^N \setminus \mathcal{A}_{[X|\Theta]}^N(\boldsymbol{\varphi}^N). \quad (110)$$

On the other hand,

$$\begin{aligned}
P_{\mathcal{A}_{12}} &= \sum_{\varphi^N \in \mathbb{F}_Q^N \setminus \mathcal{A}_{[\Theta]_\varepsilon}^N} p(\varphi^N) \sum_{\substack{\theta^N \in \mathcal{A}_{[\Theta]_\varepsilon}^N \\ \mathbf{z}_1^N \in \mathcal{A}_{[\mathbf{x}|\Theta]_\varepsilon}^N(\theta^N)}} p(\mathbf{z}_1^N | \varphi^N) \\
&\leq \sum_{\varphi^N \in \mathbb{F}_Q^N \setminus \mathcal{A}_{[\Theta]_\varepsilon}^N} p(\varphi^N) \leq \varepsilon,
\end{aligned} \tag{111}$$

since for this part

$$\bigcup_{\theta^N \in \mathcal{A}_{[\Theta]_\varepsilon}^N} \mathcal{A}_{[\mathbf{x}|\Theta]_\varepsilon}^N(\theta^N) \subseteq \mathbb{F}_Q^N. \tag{112}$$

From (109) and (111), Lemma 4 is proved.  $\blacksquare$

Now consider the term (106),

$$\begin{aligned}
P_{\mathcal{A}_2} &= \sum_{d=1}^N \sum_{\substack{\theta^N \in \mathcal{A}_{[\Theta]_\varepsilon}^N \\ \varphi^N \in \mathbb{F}_Q^N \setminus \{\theta^N\}}} \sum_{\substack{\mathbf{x}^N \in \mathcal{A}_{[\mathbf{x}|\Theta]_\varepsilon}^N(\theta^N) \\ \mathbf{z}_1^N \in \mathbb{F}_Q^N: \|\mathbf{x}^N - \mathbf{z}_1^N\|_0 = d}} p(\varphi^N, \mathbf{z}_1^N) \cdot f(d, 0; \gamma, Q, M) \\
&\leq \sum_{d=1}^{\lfloor \beta N \rfloor} \sum_{\substack{\theta^N \in \mathcal{A}_{[\Theta]_\varepsilon}^N \\ \varphi^N \in \mathbb{F}_Q^N \setminus \{\theta^N\}}} \sum_{\mathbf{z}_1^N \in \mathbb{F}_Q^N} \sum_{\mathbf{x}^N \in \mathbb{F}_Q^N: \|\mathbf{x}^N - \mathbf{z}_1^N\|_0 = d} p(\varphi^N, \mathbf{z}_1^N) \cdot f(1, 0; \gamma, Q, M) \\
&\quad + \sum_{\substack{\theta^N \in \mathcal{A}_{[\Theta]_\varepsilon}^N \\ \mathbf{x}^N \in \mathcal{A}_{[\mathbf{x}|\Theta]_\varepsilon}^N(\theta^N)}} \sum_{\substack{\varphi^N \in \mathcal{A}_{[\Theta]_\varepsilon}^N \setminus \{\theta^N\} \\ \mathbf{z}_1^N \in \mathbb{F}_Q^N}} p(\varphi^N, \mathbf{z}_1^N) \cdot f(\lfloor \beta N \rfloor, 0; \gamma, Q, M),
\end{aligned} \tag{113}$$

which is similar to (59) in Section IV-A. For  $N$  sufficient large, the condition on  $M/N$  to ensure  $P_{\mathcal{A}_2}$  tends to zero as  $N \rightarrow \infty$  is

$$\frac{M}{N} > \frac{\mathcal{H}(\Theta, \mathbf{x}) + \varepsilon}{\log Q - \xi}, \tag{114}$$

for some  $\xi \in \mathbb{R}^+$ . Finally, we have Proposition 4 to conclude the sufficient condition for reliable recovery in the NS case.

**Proposition 4** (Sufficient condition, NS case). *In the NS case, fix an arbitrary small positive real number  $\delta$ , there exists  $\varepsilon \in \mathbb{R}^+$ ,  $\xi \in \mathbb{R}^+$ ,  $N_\delta \in \mathbb{N}^+$  and  $M_\varepsilon \in \mathbb{N}^+$  such that*

for any  $N > N_\delta$  and  $M > M_\varepsilon$ , one has  $P_e < \delta$  under MAP decoding if (107) and (114) hold. One can make both  $\varepsilon$  and  $\xi$  arbitrary close to 0 as  $N \rightarrow \infty$ .

Finally, the NCS case, accounting for both communication and sensing noise, has to be considered.

**Proposition 5** (Sufficient condition, NCS case). *Considering both communication noise and sensing noise, for  $N$  and  $M$  sufficient large and positive  $\varepsilon$ ,  $\xi$  arbitrary small, the reliable recovery can be ensured under MAP decoding if*

- *the communication noise is not uniformly distributed, (65)*
- *there is no overlapping between any two different weakly conditional typical sets, i.e.,  $\mathcal{A}_{[\mathbf{x}|\Theta]^\varepsilon}^N(\boldsymbol{\theta}^N) \cap \mathcal{A}_{[\mathbf{x}|\Theta]^\varepsilon}^N(\boldsymbol{\varphi}^N) = \emptyset$  for any two typical but different  $\boldsymbol{\theta}^N$  and  $\boldsymbol{\varphi}^N$ ,*
- *the sparsity factor satisfies the constraint in (66),*
- *the compression ratio  $M/N$  is lower bounded by*

$$\frac{M}{N} > \frac{\mathcal{H}(\Theta, \mathbf{x}) + \varepsilon}{\log Q - H(p_u) - \xi}, \quad (115)$$

The derivations are similar to those of Proposition 3 and Proposition 4.

### B. Discussion and Numerical Results

When comparing the necessary condition in Proposition 1 and the sufficient condition in Proposition 5, an interesting fact is that  $\mathcal{H}(\Theta | \mathbf{x}) = 0$  is a sufficient condition to have (107). This implies that the value of  $\boldsymbol{\theta}^N$  should be fixed almost surely, as long as  $\mathbf{x}^N$  is known. So, (107) is helpful to interpret (12), justifying the need for the conditional entropy  $\mathcal{H}(\Theta | \mathbf{x})$  to tend to zero as  $N$  increases. This condition may be satisfied since  $|\mathcal{A}_{[\Theta]^\varepsilon}^N| \ll |\mathbb{F}_Q^N|$  as long as  $\mathcal{H}(\Theta) < \log Q$ . The entropy rate  $\mathcal{H}(\Theta)$  can be very small, Appendix B presents a possible situation where  $\mathcal{H}(\Theta) = 0$ . Another implicit constraint

resulting from (107) is

$$\sum_{\boldsymbol{\theta}^N \in \mathcal{A}_{[\Theta]_\varepsilon}^N} \mathcal{A}_{[\mathbf{x}|\Theta]_\varepsilon}^N(\boldsymbol{\theta}^N) \leq |\mathbb{F}_Q^N| \quad (116)$$

which means that

$$\mathcal{H}(\Theta, \mathbf{x}) \leq \log Q. \quad (117)$$

Consider a communication noise with  $\Pr(u \neq 0) = 0.1$  and the transition pmf

$$p(x_n | \theta_n) = \begin{cases} 1 - \Pr(\mathbf{x} \neq \Theta) & \text{if } x_n = \theta_n \\ \frac{\Pr(\mathbf{x} \neq \Theta)}{Q-1} & \text{if } x_n \in \mathbb{F}_Q^N \setminus \{\theta_n\} \end{cases}, \quad (118)$$

where  $\Pr(\mathbf{x} \neq \Theta)$  denotes the probability of the sensing error. In Figure 4, the lower bound of  $M/N$  is represented as a function of  $\mathcal{H}(\Theta) / \log Q$ , for different values of  $Q$  and for different values of  $\Pr(\mathbf{x} \neq \Theta)$ .

## VI. CONCLUSIONS AND FUTURE WORK

In this paper we have considered robust Bayesian compressed sensing over finite fields under MAP decoding. Both asymptotically necessary and sufficient conditions of the compression ratio for reliable recovery are obtained and their convergence is also shown, even in the case of sparse sensing matrices. Several previous results have been generalized by considering a stationary and ergodic source model. Both communication noise and sensing noise have been taken into account. We have shown that the choice of the sparsity factor of the sensing matrix only depends on the communication noise. Since necessary and sufficient conditions asymptotically converge, the MAP decoder achieves the optimum lower bound of the compression ratio, which can be expressed as a function of  $\mathcal{H}(\Theta, \mathbf{x})$ ,  $H(p_u)$ , and the alphabet size.

In this paper, the sensing matrix was assumed to be perfectly known, without specific structure. In sensor network compressive sensing applications, the structure of the sensing matrix usually depends on the structure of the network. Evaluating the impact of these

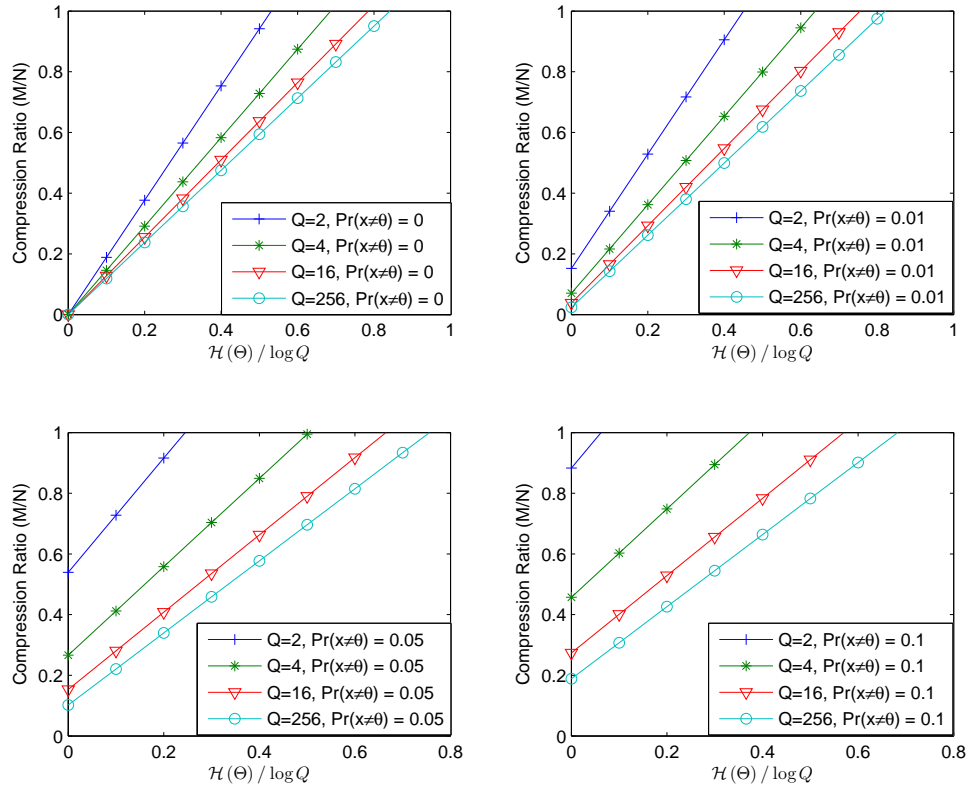


Figure 4. Optimum achievable compression ratio in function of  $\mathcal{H}(\Theta) / \log Q$ , according to (115), for the cases that  $\Pr(x \neq \Theta)$  being 0 (NC case), 0.01, 0.05, and 0.1, respectively, when  $\Pr(u \neq 0) = 0.1$

constraints on the compression efficiency will be the subject of future research. A first step in this direction was done in [15], which considered clustered sensors.

## APPENDIX A

## PROOF OF LEMMA 1

*Proof:* Let  $\mathbf{A}_i$  be the  $i$ -th row of  $\mathbf{A}$ . As all entries in  $\mathbf{A}$  are independent

$$\Pr \{ \mathbf{A} \boldsymbol{\mu}^N = \mathbf{s}^M \mid \boldsymbol{\mu}^N \neq \mathbf{0}, \mathbf{s}^M \} = \prod_{i=1}^M \Pr \{ \mathbf{A}_i \boldsymbol{\mu}^N = s_i \mid \boldsymbol{\mu}^N \neq \mathbf{0}, s_i \}. \quad (119)$$

According to [27, Lemma 21], we have

$$\Pr \{ \mathbf{A}_i \boldsymbol{\mu}^N = 0 \mid \|\boldsymbol{\mu}^N\|_0 = d_1 \} = Q^{-1} + \left( 1 - \frac{\gamma}{1 - Q^{-1}} \right)^{d_1} (1 - Q^{-1}), \quad (120)$$

and

$$\Pr \{ \mathbf{A}_i \boldsymbol{\mu}^N = q \mid \|\boldsymbol{\mu}^N\|_0 = d_1, q \in \mathbb{F}_Q \setminus \{0\} \} = Q^{-1} - \left( 1 - \frac{\gamma}{1 - Q^{-1}} \right)^{d_1} Q^{-1}. \quad (121)$$

Since  $d_2$  is the number of non-zero entries of  $\mathbf{s}^M$ , combining (119), (120), and (121), one gets

$$\begin{aligned} f(d_1, d_2; \gamma, Q, M) &= \\ &\left( Q^{-1} + \left( 1 - \frac{\gamma}{1 - Q^{-1}} \right)^{d_1} (1 - Q^{-1}) \right)^{M-d_2} \left( Q^{-1} - \left( 1 - \frac{\gamma}{1 - Q^{-1}} \right)^{d_1} Q^{-1} \right)^{d_2} \end{aligned} \quad (122)$$

The monotonicity of this function is not hard to obtain with its expression and the condition (5). ■

## APPENDIX B

A POSSIBLE SITUATION FOR  $\mathcal{H}(\Theta) = 0$ 

Consider  $N$  sensors uniformly deployed over a unit-radius disk. The physical quantities (in  $\mathbb{R}$ ), which are collected by the sensors, are denoted by  $\boldsymbol{\Omega}^N \in \mathbb{R}^N$ . We assume that  $\boldsymbol{\Omega}^N \sim \mathcal{N}(0, \boldsymbol{\Sigma})$  with

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & e^{-\lambda d_{1,2}^2} & \dots & e^{-\lambda d_{1,N}^2} \\ e^{-\lambda d_{2,1}^2} & 1 & & e^{-\lambda d_{2,N}^2} \\ \vdots & & \ddots & \vdots \\ e^{-\lambda d_{N,1}^2} & \dots & \dots & 1 \end{bmatrix}, \quad (123)$$

where  $\lambda$  is some constant,  $d_{i,j}$  is the distance between sensors  $i$  and  $j$ . The distance between two sensors is random since the location of each sensor is random. The real-valued entries of  $\mathbf{\Omega}^N$  are quantized with a  $Q$ -level scalar quantizer. We assume that  $Q = 2$ , corresponding to the rule

$$\Theta_i = \begin{cases} 0 & \text{if } \Omega_i < 0, \\ 1 & \text{if } \Omega_i \geq 0. \end{cases} \quad (124)$$

With the above assumptions, we can prove the following lemma.

**Lemma 5.** *The conditional entropy  $H(\Theta_n | \mathbf{\Theta}_1^{n-1})$  converges to zero for  $n \rightarrow \infty$ .*

*Proof:* Suppose that  $j$  is the index of the sensor which has the minimum distance to sensor  $n$ , among the  $n - 1$  neighbor sensors, i.e.,

$$j = \arg \min_{1 \leq i \leq n-1} d_{n,i}. \quad (125)$$

We have

$$H(\Theta_n | \mathbf{\Theta}_1^{n-1}) \leq H(\Theta_n | \Theta_j) \quad (126)$$

Denote the minimum distance as  $\underline{d}(n) = d_{n,j}$ , the covariance matrix of  $\Omega_n$  and  $\Omega_j$  is

$$\mathbf{\Sigma}_n = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad (127)$$

where  $\rho = e^{-\lambda \underline{d}(n)^2}$ . For a pair of realizations  $\omega_n$  and  $\omega_j$ , the joint probability density function writes

$$g(\omega_n, \omega_j) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{\omega_n^2 + \omega_j^2 - 2\rho\omega_n\omega_j}{2(1-\rho^2)}\right). \quad (128)$$

We easily obtain the probability of both  $\Omega_n$  and  $\Omega_j$  being negative,

$$\begin{aligned} \Pr\{\Omega_n < 0 \text{ and } \Omega_j < 0\} &= \int_{-\infty}^0 \int_{-\infty}^0 g(\omega_n, \omega_j) d\omega_n d\omega_j \\ &= \frac{1}{4} + \frac{1}{2\pi} \arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right) \end{aligned} \quad (129)$$

Taking into account (124), (129) is exactly the probability of the pair  $(\Theta_n, \Theta_j)$  being  $(0, 0)$ . Define

$$\varepsilon(\rho) := \frac{1}{4} - \frac{1}{2\pi} \arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right). \quad (130)$$

After the similar derivations, one obtains

$$\Pr(\Theta_n = 0, \Theta_j = 0) = \Pr(\Theta_n = 1, \Theta_j = 1) = \frac{1}{2} - \varepsilon(\rho) \quad (131)$$

and

$$\Pr(\Theta_n = 0, \Theta_j = 1) = \Pr(\Theta_n = 1, \Theta_j = 0) = \varepsilon(\rho). \quad (132)$$

Then the joint entropy is

$$H(\Theta_n, \Theta_j) = 1 + H_2(2\varepsilon(\rho)). \quad (133)$$

Meanwhile  $H(\Theta_j) = 1$ , thanks to the 2-level uniform quantizer. Obviously

$$H(\Theta_n | \Theta_j) = H_2(2\varepsilon(\rho)) = H_2\left(2\varepsilon\left(e^{-\lambda \underline{d}^2}\right)\right). \quad (134)$$

This conditional entropy is increasing in  $\underline{d}$ . When the number of sensors increases, the disk will be denser, and the minimum distance  $\underline{d}$  goes smaller. Thus,  $\underline{d}$  tends to 0 as  $n \rightarrow \infty$ , which implies that  $H(\Theta_n | \Theta_j) \rightarrow 0$ . According to (126), we conclude that  $H(\Theta_n | \Theta_1^{n-1})$  also goes to zero as  $n \rightarrow \infty$ .  $\blacksquare$

Applying the chain rule, the entropy rate writes

$$\mathcal{H}(\Theta) = \lim_{N \rightarrow \infty} \frac{H(\Theta_1) + \sum_{n=2}^N H(\Theta_n | \Theta_1^{n-1})}{N}. \quad (135)$$

By Cesaro mean [23, Theorem 4.2.3],  $\mathcal{H}(\Theta) = 0$  as  $H(\Theta_n | \Theta_1^{n-1}) \rightarrow 0$ .



## REFERENCES

- [1] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489 – 509, 2006.
- [2] E. Candes and T. Tao, “Near optimal signal recovery from random projections: Universal encoding strategies?” *IEEE Transactions on Information Theory*, vol. 52, pp. 5406 – 5425, 2006.
- [3] D. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.
- [4] E. Candes and T. Tao, “Decoding by linear programming,” *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203 – 4215, 2005.
- [5] J. Haupt and R. Nowak, “Signal reconstruction from noisy random projections,” *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4036 – 4048, 2006.
- [6] S. Ji, Y. Xue, and L. Carin, “Bayesian compressive sensing,” *IEEE Transactions on Signal Processing*, vol. 52, no. 6, pp. 2346 – 2356, 2008.
- [7] A. Montanari, “Graphical models concepts in compressed sensing,” in *Compressed Sensing: Theory and Applications*, 2012, pp. 394–438.
- [8] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 2001.
- [9] D. Baron, S. Sarvotham, and R. G. Baraniuk, “Bayesian compressive sensing via belief propagation,” *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 269 – 280, 2010.
- [10] M. Bayati and A. Montanari, “The dynamics of message passing on dense graphs, with applications to compressed sensing,” *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764– 785, 2011.
- [11] I. Akyildi, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: a survey,” *Computer Networks*, vol. 38, pp. 393–422, 2002.
- [12] T. Ho, M. Medard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, “A random linear network coding approach to multicast,” *IEEE Transactions on Information Theory*, vol. 52, pp. 4413–4430, 2006.
- [13] L. Iwaza, M. Kieffer, and K. Al-Agha, “Map estimation of network-coded correlated sources,” *Proc. of ATC*, vol. Hanoi, Vietnam, 2012.
- [14] F. Bassi, C. Liu, L. Iwaza, and M. Kieffer, “Compressive linear network coding for efficient data collection in wireless sensor networks,” in *Proc. 20th European Signal Processing Conf. (EUSIPCO)*, Bucharest, Romania, 2012, pp. 714 – 718.
- [15] K. Rajawat, C. Alfonso, and G. Giannakis, “Network-compressive coding for wireless sensors with correlated data,” *IEEE Transactions on Communications*, vol. 11, no. 12, pp. 4264–4274, 2012.
- [16] I. Bourtsoulatzé, N. Thomos, and P. Frossard, “Correlation-aware reconstruction of network coded sources,” in *Proc. IEEE International Symposium on Network Coding (NetCod)*, June 2012, pp. 91–96, Cambridge, MA.

- [17] M. Nabeae and F. Labeau, “Restricted isometry property in quantized network coding of sparse messages,” in *IEEE Global Communications Conference (GLOBECOM)*, 2012, pp. 112–117.
- [18] S. C. Draper and S. Malekpour, “Compressed sensing over finite fields,” in *Proc. IEEE Intl. Symp. on Info. Theory (ISIT)*, Seoul, Korea, 2009, pp. 669 – 673.
- [19] J.-T. Seong and H.-N. Lee, “Necessary and sufficient conditions for recovery of sparse signals over finite fields,” *IEEE Communications Letters*, vol. 17, no. 10, pp. 1976 – 1979, 2013.
- [20] I. Csiszar, “The method of types,” *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505–2523, 1998.
- [21] P. Chou, Y. Wu, and K. Jain, “Practical network coding,” *Proc. of the 41-st Allerton Conference*, vol. Monticello, IL, 2003.
- [22] M. Jafari, L. Keller, C. Fragouli, and K. Argyraki, “Compressed network coding vectors,” in *Proc. IEEE Intl. Symp. on Info. Theory (ISIT)*, Seoul, Korea, 2009, pp. 109–113.
- [23] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.
- [24] I. Csiszar, “Linear codes for sources and source networks: Error exponents, universal coding,” *IEEE Transactions on Information Theory*, vol. 28, no. 4, pp. 585–592, 1982.
- [25] R. Yeung, “A first course in information theory,” MA: *Kluwer*, vol. Princeton, NJ, 2004.
- [26] R. Gallager, “Information theory and reliable communication,” *John Wiley and Sons.*, 1968.
- [27] V. Tan, L. Balzano, and S. Draper, “Rank minimization over finite fields: fundamental limits and coding-theoretic interpretations,” *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 2018–2039, 2012.